

STIC-Biotech/ChemLib

78067

From: Rao, Manjunath N.
Sent: Thursday, October 17, 2002 12:57 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/687860

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room ~~10C-01~~
Phone: 306-5681

→ 100001

CRFF

Date: 10-17-02

Please search the following as soon as possible for application with serial number **09/687,860**

SEQ ID NO:24, against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If y u have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Requestor's

Name: _____

Serial

Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

BEST AVAILABLE COPY

STAFF USE ONLY

Date completed: 10-18-02

Searcher: Beverly e 4994

Terminal time: 20

Elapsed time: _____

CPU time: _____

Total time: 23

Number of Searches: _____

Number of Databases: 1

Search Site

_____ STIC

_____ CM-1

_____ Pre-S

Type of Search

_____ N.A. Sequence

_____ A.A. Sequence

_____ Structure

_____ Bibliographic

Vendors

_____ IG Suite

_____ STN

_____ Dialog

_____ APS

_____ Geninfo

_____ SDC

_____ DARC/Questel

_____ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 15:52:13 ; Search time 35 seconds
(without alignments)
1754.966 Million cell updates/sec

Title: US-09-687-860-24

Perfect score: 3060

Sequence: 1 MRLPMSIALPLLSWVAGF.....VDGLVLSGLCPDLSLVDD 553

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3060	100.0	553	AA801423	Human TANGO 212.
2	3055	99.8	553	AA893622	Human polypeptide,
3	3049.5	99.7	554	AA893156	Human polypeptide,
4	3049.5	99.7	554	AA872224	Human EXMAD-2 SEQ
5	3030	99.0	553	AA18108	Protein encoded by
6	2986.5	97.6	573	AA840942	Human polypeptide
7	2767	90.4	537	AA18110	Protein encoded by
8	1931	63.1	338	AA11702	Human PRO320 prote
9	1931	63.1	338	AA844258	Human PRO320 (UM2
10	1931	63.1	338	AA18669	Amino acid sequenc
11	1931	63.1	338	AA195339	Human PRO320 anti

12	1894	61.9	331	22	AA823677	Human EST encoded
13	1491	48.7	333	21	AA801429	Murine TANGO 212.
14	1288	42.1	284	21	AA176110	Rat TGF-beta homol
15	1288	42.1	284	22	AA856049	Skin cell protein,
16	1105	36.1	546	22	AA870549	Clone 1646/945.0.8
17	1091.5	35.7	582	22	AA870547	Human PRO17 protei
18	1030	33.7	509	20	AA13397	Amino acid sequenc
19	1030	33.7	509	22	AA129049	Human PRO polypept
20	1030	33.7	509	22	AA880265	Human PRO334 prote
21	976	31.9	251	22	AB10214	Human cDNA SEQ ID
22	976	31.9	251	22	AA16935	Human novel secret
23	683	22.3	183	21	AA842085	Human ORFX ORF1849
24	661	21.6	211	22	AA004843	Human polypeptide
25	602	19.7	100	20	AA18109	EGF motif containi
26	409.5	13.4	2912	22	AB806402	Novel human diagno
27	393	12.8	1118	22	AA850209	Human fibrillin-11
28	372.5	12.2	2189	11	AA805222	Human fibrillin-11
29	369.5	12.1	333	21	AA84709	Antigen GX5401FL e
30	367.5	12.0	636	22	AA840146	Amino acid sequenc
31	362.5	11.8	576	22	AA880174	Human polypeptide
32	362.5	11.8	576	22	AA880175	Human protein SEQ
33	362.5	11.8	576	22	AA841932	Human polypeptide
34	362.5	11.8	603	22	AA879191	Human protein SEQ
35	362.5	11.8	650	22	AA879190	Human protein SEQ
36	357	11.7	1121	22	AA878887	Human fibrillin 3.
37	354.5	11.6	1833	16	AA879478	Mouse LTRP-2. Mus
38	354.5	11.6	1833	21	AA812270	Human latent TGF-b
39	350.5	11.5	497	22	AA850207	Human fibrillin-11
40	350	11.4	683	12	AA811150	Human fibrillin-11
41	350	11.4	683	18	AA827600	Fibulin C. Homo s
42	350	11.4	686	22	AA819385	Human fibrulin type
43	348	11.4	585	22	AA867242	Novel human diagno
44	339.5	11.1	741	22	AA895002	Amino acid sequenc
45	339.5	11.1	780	22	AA867241	Human protein sequ
						Amino acid sequenc

ALIGNMENTS

RESULT 1	AA801423	standard; Protein: 553 AA.
ID	AA801423	
XX	AA801423:	
AC	20-OCT-2000	(first entry)
XX		
DE	Human TANGO 212.	
XX		
DE	TANGO: 128: 140: 197: 212: 213: 224: 239: modulating agent; asthma;	
KW	graft versus-host diseases; rheumatoid arthritis; psoriasis;	
KW	inflammatory bowel disease; septic shock; ulcerative colitis;	
KW	Crohn's disease; chronic myelogenous leukemia; cancer; liver	
KW	disease; Hodgkin's disease; osteoarthritis; Lyme's disease;	
KW	cachexia; autoimmune disease; myasthenia gravis; autoimmune	
KW	systemic lupus erythematosus; transgenic animal; diagnosis;	
KW	prognosis; prophylactic; therapeutic; human.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200039284-A1.	
XX		
PD	06-JUL-2000.	
XX		
PF	23-DEC-1999: 99WO-US31025.	
XX		
PR	30-DEC-1998: 98US-0223546.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		
PI	Holtzman DA:	
XX		
DR	WPI: 2000-465743/40.	

DR N-PSDB; AAA47456.
 XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
 PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
 PT arthritis, psoriasis and autoimmune diseases
 XX
 PS Claim 8; Fig 5; 209pp; English.
 XX
 CC Nucleic acids encoding TANGO polypeptides are useful as modulating
 CC agents for regulating cellular processes like asthma, graft
 CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
 CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,
 CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
 CC lupus erythematosus. The nucleic acids are also useful for producing
 CC transgenic animals and the TANGO polypeptides themselves. Partial
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
 CC forensic biology, for diagnostic assays, prognostic assays,
 CC pharmacogenomics and for monitoring clinical trials. TANGO
 CC polypeptides are suitable for both prophylactic and therapeutic
 CC methods for treating a subject at risk of a disorder or having a
 CC disorder associated with aberrant TANGO expression. A wide range
 CC of cellular disorders can be treated.
 CC
 XX
 XX Sequence 553 AA:

Query Match 100.0%; Score 3060; DB 21; Length 553;
 Best Local Similarity 100.0%; Pred. No. 1.8e-217;
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPMSIALPRLLSVWAGFGNMSARHGHGLASARQGVCHYGTKLACCGWRNSKGV 60
 DB 1 MPLPMSIALPRLLSVWAGFGNMSARHGHGLASARQGVCHYGTKLACCGWRNSKGV 60
 QY 61 CEATEPCGKFGECYGPVKRCFPGYTGKTCSDQVNECGMRRPQOHNCVMTGSKFC 120
 DB 61 CEATEPCGKFGECYGPVKRCFPGYTGKTCSDQVNECGMRRPQOHNCVMTGSKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLPSSGLRLAPNGRDLIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLPSSGLRLAPNGRDLIDECAS 180
 QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELYISGRDCTDINECTYDSTCHSHANCFMTQ 240
 DB 181 GKVICPYNRRCVNTFGSYCKCHIGFELYISGRDCTDINECTYDSTCHSHANCFMTQ 240
 QY 241 GSFCKCKOGYKGNLGRCSAIPENSVKEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT 300
 DB 241 GSFCKCKOGYKGNLGRCSAIPENSVKEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT 300
 QY 301 PEPTPTPPKYNLOPFNTEEIVSRGNSHGKKGNEKKMEGLDEKREERALKNDIEER 360
 DB 301 PEPTPTPPKYNLOPFNTEEIVSRGNSHGKKGNEKKMEGLDEKREERALKNDIEER 360
 QY 361 SLRGVFFPKVNEGEFLLIVORKALSKLEHKDINTSVCSFNGHICDMKODEDEDFD 420
 DB 361 SLRGVFFPKVNEGEFLLIVORKALSKLEHKDINTSVCSFNGHICDMKODEDEDFD 420
 QY 421 WNPADRDNAIGFYMAVPLAGHKDGRKLLPDLQPOSNCLFDFRLAGDKGKRLV 480
 DB 421 WNPADRDNAIGFYMAVPLAGHKDGRKLLPDLQPOSNCLFDFRLAGDKGKRLV 480
 QY 481 FVKNNSNNAIAEKTTSSEDEKMTGKIQLYOGTDATKSIIFEAERKGTGEIADVGLLV 540
 DB 481 FVKNNSNNAIAEKTTSSEDEKMTGKIQLYOGTDATKSIIFEAERKGTGEIADVGLLV 540
 QY 541 SGICPDSLVSVD 553
 DB 541 SGICPDSLVSVD 553

RESULT 2

AAM93622
 ID AAM93622 standard; Protein; 553 AA.
 XX
 AC AAM93622;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 3456.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 XX
 PR 11-JAN-2000; 2000JP-0118774.
 XX
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI: 2001-524255/58.
 XX N-PSDB; AAK94555.
 XX

830 Primers useful for synthesizing full length cDNA clones and their
 use in genetic manipulation -

Claim 8; SEQ ID NO 3456; 1380bp + sequence listing; English.

The invention relates to primers for synthesizing full length cDNA
 clones. 830 cDNA molecules encoding a human protein have been
 isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 molecules have been determined. Primers for synthesizing the full length
 cDNA are useful for clarifying the function of the protein encoded by
 the cDNA. The full length clones were obtained by construction of full
 length enriched cDNA libraries that were synthesised by the oligo-capping
 method. The primers enable the production of the full length cDNA easily
 without any special methods. The present sequence is a polypeptide
 encoded by a full length human cDNA of the invention.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in CD-ROM format directly from EPO.

Sequence 553 AA:

Query Match 99.8%; Score 3055; DB 22; Length 553;
 Best Local Similarity 99.8%; Pred. No. 4.2e-217;
 Matches 552; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPMSIALPRLLSVWAGFGNMSARHGHGLASARQGVCHYGTKLACCGWRNSKGV 60
 DB 1 MPLPMSIALPRLLSVWAGFGNMSARHGHGLASARQGVCHYGTKLACCGWRNSKGV 60
 QY 61 CEATEPCGKFGECYGPVKRCFPGYTGKTCSDQVNECGMRRPQOHNCVMTGSKFC 120
 DB 61 CEATEPCGKFGECYGPVKRCFPGYTGKTCSDQVNECGMRRPQOHNCVMTGSKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLPSSGLRLAPNGRDLIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEBGPQCLPSSGLRLAPNGRDLIDECAS 180
 QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELYISGRDCTDINECTYDSTCHSHANCFMTQ 240
 DB 181 GKVICPYNRRCVNTFGSYCKCHIGFELYISGRDCTDINECTYDSTCHSHANCFMTQ 240
 QY 241 GSFCKCKOGYKGNLGRCSAIPENSVKEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT 300
 DB 241 GSFCKCKOGYKGNLGRCSAIPENSVKEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT 300

```

Db 241 GSEFKCKOGYKGNGLRCSAIPENSVEVLRAPTIKRIKLLAHKNSMKKAKIKNVY 300
QY 301 PEPTRTPTPKVNIQPFNYEEIVSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNIQPFNYEEIVSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
QY 361 SLRGDVEFFPKYNEAGEGGLILVORKALTSKLEHKNLNIYVDCSFNHCICWKODREDFD 420
Db 361 SLRGDVEFFPKYNEAGEGGLILVORKALTSKLEHKNLNIYVDCSFNHCICWKODREDFD 420
QY 421 WNPADRDNAIGFYMAVPALAGHKKDIGRLKLLPDLPQSNFCLLPYRLAGDKVGLRV 480
Db 421 WNPADRDNAIGFYMAVPALAGHKKDIGRLKLLPDLPQSNFCLLPYRLAGDKVGLRV 480
QY 481 FVKNNSNNALAMEKTTSEDEKMTGKIQLYOGTDATKSIIFEAERGKGTGEIADVGLLV 540
Db 481 FVKNNSNNALAMEKTTSEDEKMTGKIQLYOGTDATKSIIFEAERGKGTGEIADVGLLV 540
QY 541 SGLCPDLSLVDD 553
Db 541 SGLCPDLSLVDD 553
QY 541 SGLCPDLSLVDD 553
Db 541 SGLCPDLSLVDD 553

RESULT 3
AAM39156
ID AAM39156 standard; Protein; 554 AA.
XX
AC AAM39156;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2301.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA158312.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2301; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

```

```

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 554 AA;
QY Query Match 99.7%; Score 3049.5; DB 22; Length 554;
Best Local Similarity 99.8%; Pred. No. 1.1e-216;
Matches 553; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 MRLPMSIALPRLLSWVAGFENASARRHGLLASARQGVCHYGTAKLACCGWRNRSKV 60
Db 1 MRLPMSIALPRLLSWVAGFENASARRHGLLASARQGVCHYGTAKLACCGWRNRSKV 60
QY 61 CEATCEPGCKRGECVGPNCRCRPGYTGKTCSDPVNECGMPPRCQHRVTHGSYKFC 120
Db 61 CEATCEPGCKRGECVGPNCRCRPGYTGKTCSDPVNECGMPPRCQHRVTHGSYKFC 120
QY 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPOCLCPSSGLRLAPNGRDLIDECAS 180
Db 121 LSGHMLMPDATCVNSRTCAMINCOYSCEDTEEGPOCLCPSSGLRLAPNGRDLIDECAS 180
QY 181 GKVICPTNRCVNTFSGSYCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
Db 181 GKVICPTNRCVNTFSGSYCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTQ 240
QY 241 GSEFKCKOGYKGNGLRCSAIPENSVEVLRAPTIKRIKLLAHKNSMKKAKIKNVY 300
Db 241 GSEFKCKOGYKGNGLRCSAIPENSVEVLRAPTIKRIKLLAHKNSMKKAKIKNVY 300
QY 301 PEPTRTPTPKVNIQPFNYEEIVSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
Db 301 PEPTRTPTPKVNIQPFNYEEIVSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
QY 361 SLRGDVEFFPKYNEAGEGGLILVORKALTSKLEHKNLNIYVDCSFNHCICWKODREDFD 420
Db 361 SLRGDVEFFPKYNEAGEGGLILVORKALTSKLEHKNLNIYVDCSFNHCICWKODREDFD 420
QY 420 WNPADRDNAIGFYMAVPALAGHKKDIGRLKLLPDLPQSNFCLLPYRLAGDKVGLRV 479
Db 420 WNPADRDNAIGFYMAVPALAGHKKDIGRLKLLPDLPQSNFCLLPYRLAGDKVGLRV 479
QY 481 FVKNNSNNALAMEKTTSEDEKMTGKIQLYOGTDATKSIIFEAERGKGTGEIADVGLLV 539
Db 481 FVKNNSNNALAMEKTTSEDEKMTGKIQLYOGTDATKSIIFEAERGKGTGEIADVGLLV 539
QY 540 VSGLCPSDLSLVDD 553
Db 540 VSGLCPSDLSLVDD 553

RESULT 4
AAB27224
ID AAB27224 standard; Protein; 554 AA.
XX
AC AAB27224;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human EXMAD-2 SEQ ID NO: 2.
XX
KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;

```

KM Inflammation; reproductive disorder; cardiovascular disorder;
 KM Immune disorder; musculoskeletal disorder; developmental disorder;
 KM Gastrointestinal disorder; cell proliferation disorder.
 OS Homo sapiens.
 XX MO200068380-A2.
 PN 16-NOV-2000.
 PD 10-MAY-2000; 2000MO-US12811.
 PF 11-MAY-1999; 99US-0133643.
 PR 23-AUG-1999; 99US-0150409.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Bandman O, Hillman JL, Tang YT, Lai P, Yue H, Baughn MR, Lu DM;
 PI Azimzai Y;
 XX WPI: 2001-007395/01.
 DR N-PSDB; AAC66891.
 XX Claim 1; Page 88-89; 129pp; English.
 PS Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 XX
 CC The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 CC
 SO Sequence 554 AA;
 Query Match 99.7%; Score 3049.5; DB 22; Length 554;
 Best Local Similarity 99.8%; Pred. No. 1.1e-216;
 Matches 553; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 MPLPMSIALPLLISWVAGFGNMAASARHHGLASAROPGVCHYGKTLACCYGMRNSKV 60
 DB 1 MPLPMSIALPLLISWVAGFGNMAASARHHGLASAROPGVCHYGKTLACCYGMRNSKV 60
 QY 61 CEATCEPCRCFEGCVGNKRCRCPFGYTGKTCSDVNECGMKPRPOHRCVNTHSGYKFC 120
 DB 61 CEATCEPCRCFEGCVGNKRCRCPFGYTGKTCSDVNECGMKPRPOHRCVNTHSGYKFC 120
 QY 121 LSGHMLMPDARCVNSRRCAMINCOYSCEDTREGQCICPSSGRLAPRGRCCLDIDECAS 180
 DB 121 LSGHMLMPDARCVNSRRCAMINCOYSCEDTREGQCICPSSGRLAPRGRCCLDIDECAS 180
 QY 181 GKVICPNRRCVNTEFGSYKCHIGFELYISGRYDCIDINECTMDSTGSHNACFTQ 240
 DB 181 GKVICPNRRCVNTEFGSYKCHIGFELYISGRYDCIDINECTMDSTGSHNACFTQ 240
 QY 241 GSFCKCKOGYKGNLRCASAIPENSYKVELRAPGTIKDIRIKLLAHKNSMKKAKIKKAVT 300
 DB 241 GSFCKCKOGYKGNLRCASAIPENSYKVELRAPGTIKDIRIKLLAHKNSMKKAKIKKAVT 300
 QY 301 PEPTPTTPKYNVLOPNTVEEIVSRGNSHGKKGNEEKMKEGLEDEKREKALKNDIEER 360
 DB 301 PEPTPTTPKYNVLOPNTVEEIVSRGNSHGKKGNEEKMKEGLEDEKREKALKNDIEER 360
 QY 361 SLRGDVFEPKYNAGEGGLIYORKALTSKLEHK-DLNTISVDCSPFNHICGMKODREDF 419
 DB 361 SLRGDVFEPKYNAGEGGLIYORKALTSKLEHKADLNTISVDCSPFNHICGMKODREDF 420

QY 420 DWNPADRDNAIGFYMAVPALAGHKKIDIRKLPLDLPDSNFCLLFPYRLAGDKVGLR 479
 DB 421 DWNPADRDNAIGFYMAVPALAGHKKIDIRKLPLDLPDSNFCLLFPYRLAGDKVGLR 480
 QY 480 VEVKNSNNAWAMEKTTSEDEKWKTKGIQLYOGDTATKSIIFEAEKRGKGTGEIADVGLL 539
 DB 481 VEVKNSNNAWAMEKTTSEDEKWKTKGIQLYOGDTATKSIIFEAEKRGKGTGEIADVGLL 540
 QY 540 VSGLCPSDLSVDD 553
 DB 541 VSGLCPSDLSVDD 554
 -RESULT 5
 ID AAY18108 standard; Protein: 553 AA.
 AC AAY18108;
 XX 10-AUG-1999 (first entry)
 DE Protein encoded by cDNA insert of clone pEGFR-HY2.
 XX
 KW Epidermal growth factor; EGF repeat domain; haematopoiesis regulator;
 KW tissue growth activity; activin; inhibin; chemotaxis; chemokinesis;
 KW haemostasis; thrombolysis; anti-inflammatory; leukaemia; anaemia;
 KW immune disorder; immune deficiency; nervous system disorder; therapy.
 XX
 OS Synthetic.
 PN WO9927096-A1.
 XX 03-JUN-1999.
 PD 23-NOV-1998; 98WO-US24524.
 PP 22-NOV-1997; 97US-0968800.
 PR (HYSE-) HYSEQ INC.
 PA Crivenjakov R, Dickson M, Drmanac RT, Drmanac S;
 PI Ford J, Kita D, Labat I, Leshkowitz D;
 DR WPI: 1999-370904/31.
 DR N-PSDB; AAX79501.
 XX New polypeptide with epidermal growth factor repeat domains
 PS Claim 8; Fig 5; 96pp; English.
 XX This sequence represents a polypeptide of the invention, which has
 CC similarity to epidermal growth factor (EGF) repeat domains. The
 CC polypeptides and their compositions may have haematopoiesis regulating,
 CC tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic,
 CC thrombolytic, receptor/ligand and anti-inflammatory activities. They may
 CC be used to treat leukemias, anaemias, immune disorders and deficiencies
 CC and nervous system disorders. They can be used in screening assays to
 CC identify agents which bind to them and the nucleotide sequences can be
 CC used as probes for in situ hybridisation. The polypeptides and their
 CC polynucleotides can also be used for other therapeutic, diagnostic and
 CC research utilities.
 SO Sequence 553 AA;
 Query Match 99.0%; Score 3030; DB 20; Length 553;
 Best Local Similarity 99.3%; Pred. No. 2.9e-215;
 Matches 549; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MPLPMSIALPLLISWVAGFGNMAASARHHGLASAROPGVCHYGKTLACCYGMRNSKV 60
 DB 1 MPLPMSIALPLLISWVAGFGNMAASARHHGLASAROPGVCHYGKTLACCYGMRNSKV 60

Db 558 LTVSGLCPDLSLSD 573

RESULT 7
AA18110
ID AA18110 standard; Protein: 537 AA.
XX
XX AA18110;
AC
XX 10-AUG-1999 (first entry)
DT
XX
XX Protein encoded by fragment of cDNA insert of clone pEGFR-HY2.
DE
XX
XX Epidermal growth factor; EGF repeat domain; haematopoiesis regulator;
KW tissue growth activity; activin; inhibin; chemotaxis; chemokinesis;
KW haemostasis; thrombolysis; anti-inflammatory; leukaemia; anaemia;
KW immune disorder; immune deficiency; nervous system disorder; therapy.
XX
XX Synthetic.
OS
XX WO9927096-A1.
PN
XX 03-JUN-1999.
PD
XX 23-NOV-1998; 98WO-US24524.
PE
XX 22-NOV-1997; 97US-0968800.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Crivenjakov R, Dickson M, Drmanac RT, Drmanac S;
PI Ford J, Kita D, Labat I, Leshkowitz D;
XX
XX WPI: 1999-370904/31.
DR N-PSDB; AAX79503.
XX
XX New polypeptide with epidermal growth factor repeat domains
PT
XX
XX Claim 21; Fig 3; 96pp; English.
PS
XX
XX This sequence represents a polypeptide of the invention, which has
CC similarity to epidermal growth factor (EGF) repeat domains. The
CC polypeptides and their compositions may have haematopoiesis regulating,
CC tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic,
CC thrombolytic, receptor/ligand and anti-inflammatory activities. They may
CC be used to treat leukaemias, anaemias, immune disorders and deficiencies
CC and nervous system disorders. They can be used in screening assays to
CC identify agents which bind to them and the nucleotide sequences can be
CC used as probes for in situ hybridisation. The polypeptides and their
CC polynucleotides can also be used for other therapeutic, diagnostic and
CC research utilities.
XX
XX Sequence 537 AA;
SQ

Query Match 90.4%; Score 2767; DB 20; Length 537;
Best Local Similarity 99.6%; Pred. No. 7.1e-196;
Matches 500; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 52 GMRNSKGVCEATCEPGKFGECVGNPKRCPPGYTGKTSQDVNDCGKMPRPOHRCVN 111
DB 1 GMRNSKGVCEATCEPGKFGECVGNPKRCPPGYTGKTSQDVNDCGKMPRPOHRCVN 60
QY 112 THGSYKCFCLSGHMLMPATCVNSFTCAMINCQYSCDETEBGPQCLPSSGRLAPNGRD 171
DB 61 THGSYKCFCLSGHMLMPATCVNSFTCAMINCQYSCDETEBGPQCLPSSGRLAPNGRD 120
QY 172 CLIDECASGVYICPNRCVNTFESYCKCHIGFELYISGRVNCIDINETMNSHCS 231
DB 121 CLIDECASGVYICPNRCVNTFESYCKCHIGFELYISGRVNCIDINETMNSHCS 180
QY 232 HHANCFNTQGSFKCKCKOGYKNGJLCSAIPENSVKEYLRAFGTITKDRITKLAKNSMK 291
DB 181 HHANCFNTQGSFKCKCKOGYKNGJLCSAIPENSVKEYLRAFGTITKDRITKLAKNSMK 240

QY 292 KKAQIKNVTPEPTPTPEKVINQIPFNYEIRVSRGNSHGKKGNEKMEKEGLEDKREK 351
DB 241 KKAQIKNVTPEPTPTPEKVINQIPFNYEIRVSRGNSHGKKGNEKMEKEGLEDKREK 300
QY 352 ALKNDIEERSLGDVFFPKVNEAGEPGLIVORKALTSKLEHKDLNISVDCSFNHCIDW 411
DB 301 ALKNDIEERSLGDVFFPKVNEAGEPGLIVORKALTSKLEHKDLNISVDCSFNHCIDW 360
QY 412 KODREDDPDMNPADBDNALGFMAVPALAGHKDDIGRLKLLPDLOPOSFCLLPDYRLA 471
DB 361 KODREDDPDMNPADBDNALGFMAVPALAGHKDDIGRLKLLPDLOPOSFCLLPDYRLA 420
QY 472 GDKVKKLRYFVKNSSNNALAMEKTSDEKMKTKGIQLYOGTDTATSIIFEARGKGTGE 531
DB 421 GDKVKKLRYFVKNSSNNALAMEKTSDEKMKTKGIQLYOGTDTATSIIFEARGKGTGE 480
QY 532 IAVDGLVYSGLCPSLSLSD 553
DB 481 IAVDGLVYSGLCPSLSLSD 502

RESULT 8
AA1702
ID AA171702 standard; Protein: 338 AA.
XX
XX AA171702;
AC
XX
XX 07-DEC-1999 (first entry)
DT
XX
XX Human PRO320 protein sequence.
DE
XX
XX Human; PRO: EST: expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
XX Homo sapiens.
OS
XX
XX WO9946281-A2.
PN
XX
XX 16-SEP-1999.
PD
XX
XX 08-MAR-1999; 99WO-US05028.
PS
XX
XX 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 25-MAR-1998; 98US-0079656.
PR 26-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.

PR	08-APR-1998;	98US-0081070.
PR	08-APR-1998;	98US-0081071.
PR	09-APR-1998;	98US-0081195.
PR	09-APR-1998;	98US-0081203.
PR	09-APR-1998;	98US-0081229.
PR	15-APR-1998;	98US-0081817.
PR	15-APR-1998;	98US-0081838.
PR	15-APR-1998;	98US-0081952.
PR	15-APR-1998;	98US-0081955.
PR	21-APR-1998;	98US-0082568.
PR	21-APR-1998;	98US-0082569.
PR	22-APR-1998;	98US-0082700.
PR	22-APR-1998;	98US-0082704.
PR	22-APR-1998;	98US-0082804.
PR	23-APR-1998;	98US-0082767.
PR	23-APR-1998;	98US-0082796.
PR	27-APR-1998;	98US-0083336.
PR	28-APR-1998;	98US-0083342.
PR	29-APR-1998;	98US-0083392.
PR	29-APR-1998;	98US-0083495.
PR	29-APR-1998;	98US-0083496.
PR	29-APR-1998;	98US-0083439.
PR	29-APR-1998;	98US-0083500.
PR	29-APR-1998;	98US-0083545.
PR	29-APR-1998;	98US-0083554.
PR	29-APR-1998;	98US-0083558.
PR	29-APR-1998;	98US-0083559.
PR	30-APR-1998;	98US-0083742.
PR	05-MAY-1998;	98US-0084366.
PR	06-MAY-1998;	98US-0084414.
PR	06-MAY-1998;	98US-0084441.
PR	07-MAY-1998;	98US-0084598.
PR	07-MAY-1998;	98US-0084600.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084639.
PR	07-MAY-1998;	98US-0084640.
PR	07-MAY-1998;	98US-0084643.
PR	13-MAY-1998;	98US-0085343.
PR	13-MAY-1998;	98US-0085338.
PR	13-MAY-1998;	98US-0085339.
PR	15-MAY-1998;	98US-0085573.
PR	15-MAY-1998;	98US-0085579.
PR	15-MAY-1998;	98US-0085580.
PR	15-MAY-1998;	98US-0085582.
PR	15-MAY-1998;	98US-0085689.
PR	15-MAY-1998;	98US-0085697.
PR	15-MAY-1998;	98US-0085700.
PR	15-MAY-1998;	98US-0085704.
PR	18-MAY-1998;	98US-0086023.
PR	22-MAY-1998;	98US-0086392.
PR	22-MAY-1998;	98US-0086414.
PR	22-MAY-1998;	98US-0086430.
PR	22-MAY-1998;	98US-0086486.
PR	28-MAY-1998;	98US-0087098.
PR	28-MAY-1998;	98US-0087106.
PR	28-MAY-1998;	98US-0087208.
PR	30-JUL-1998;	98US-0094651.
PR	11-SEP-1998;	98US-0100038.

PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
XX WPI: 1999-551358/46.
DR N-PSDB; AA233991.
DR
XX
XX New secreted and transmembrane polypeptides and their polynucleotides
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX
PS Claim 12; Fig 45; 530pp; English.
XX

Query Match	Best Local Similarity	63.18; Score 1931; DB 20; Length 338;
Matches 338; Conservative 100.0%; Pred. 1.9e-134;	0; Mismatches 0; Indels 0; Gaps 0;	
1 MRLPMSLALPRLLSWVAGFGMAAARHHGGLASAROPGVCHYGTGKLACCYGMRNSKGV 60		
1 MRLPMSLALPRLLSWVAGFGMAAARHHGGLASAROPGVCHYGTGKLACCYGMRNSKGV 60		
61 CEATCEPCCKFECEYGVNPKRCRCFPGYTGTCSQDVNECGMRRPCONRCVNTHGSYKFCG 120		
61 CEATCEPCCKFECEYGVNPKRCRCFPGYTGTCSQDVNECGMRRPCONRCVNTHGSYKFCG 120		
121 LSGHMLMDATCVNSTRCTAMINCOYSCEDTEBGPCLCPSSGLRLAPRGRCLODIECAS 180		
121 LSGHMLMDATCVNSTRCTAMINCOYSCEDTEBGPCLCPSSGLRLAPRGRCLODIECAS 180		
181 GKVICPYNRRCVNTEGSGYYCKCHIEFELQYISGRDCCIDINECTMDSTHSCSHANCFMTQ 240		
181 GKVICPYNRRCVNTEGSGYYCKCHIEFELQYISGRDCCIDINECTMDSTHSCSHANCFMTQ 240		
241 GSFKRCCKQGYTGNLRCSSAIPENSVKEYLAPGTIKRIKLLAHKNSMKAATIKVYT 300		
241 GSFKRCCKQGYTGNLRCSSAIPENSVKEYLAPGTIKRIKLLAHKNSMKAATIKVYT 300		
301 PEPTRTPTPKVNLQPFNYEEIYSRGNSHGKKGMEK 338		
301 PEPTRTPTPKVNLQPFNYEEIYSRGNSHGKKGMEK 338		
Db		
301 PEPTRTPTPKVNLQPFNYEEIYSRGNSHGKKGMEK 338		
RESULT 9		
AAB44258		
ID AAB44258 standard; Protein: 338 AA.		
XX		
AC AAB44258;		
XX		
DT 08-FEB-2001 (first entry)		
XX		
DE Human PRO320 (UNQ281) protein sequence SEQ ID NO:119.		
XX		
KW Human: secreted protein; transmembrane protein; PRO; EST; cytosolic; expressed sequence tag; detection; cancer.		
XX		
OS Homo sapiens.		
OS		
PN WO200053756-A2.		
XX		
PD 14-SEP-2000.		
XX		
PF 18-FEB-2000; 2000WO-US04341.		
XX		
PR 08-MAR-1999; 99WO-US05028.		
PR 12-MAR-1999; 99US-0123957.		
PR 29-MAR-1999; 99US-0126773.		
PR 21-APR-1999; 99US-0130229.		
PR 28-APR-1999; 99US-0131445.		
PR 14-MAY-1999; 99US-0134287.		
PR 23-JUN-1999; 99US-0141037.		
PR 26-JUL-1999; 99US-0145698.		
PR 29-OCT-1999; 99US-0162506.		

PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gertlisen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kljavin LJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 XX WPI; 2000-611443/58.
 DR N-PSDB; AAC78484.
 XX
 XX Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 PS
 PS Claim 12; Fig 45; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytosstatic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 CC
 XX
 SQ Sequence 338 AA;
 Query Match 63.1%; Score 1931; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRLPMSIALPILLSVWAGFGNNAASARHHGLASAROPGVCHYGTKLACCYGMRNSKGV 60
 DB 1 MRLPMSIALPILLSVWAGFGNNAASARHHGLASAROPGVCHYGTKLACCYGMRNSKGV 60
 QY 61 CEATCEPCCKGECVGPYKRCRFGYTGKTSQDVNECGMKRPRCOHRCVNTHSGYKFC 120
 DB 61 CEATCEPCCKGECVGPYKRCRFGYTGKTSQDVNECGMKRPRCOHRCVNTHSGYKFC 120
 QY 121 LSGHMLMPDADCVNSRTCAMINCOVSCDETEBGPCLCPSSGLRLAPRGROCLDIDECAS 180
 DB 121 LSGHMLMPDADCVNSRTCAMINCOVSCDETEBGPCLCPSSGLRLAPRGROCLDIDECAS 180
 QY 121 LSGHMLMPDADCVNSRTCAMINCOVSCDETEBGPCLCPSSGLRLAPRGROCLDIDECAS 180
 DB 121 LSGHMLMPDADCVNSRTCAMINCOVSCDETEBGPCLCPSSGLRLAPRGROCLDIDECAS 180
 QY 181 GKVICPVNRCVNTFGSYKCHIGFELQYISGRYCDIDINECTMDSHTCSHHANCFTQ 240
 DB 181 GKVICPVNRCVNTFGSYKCHIGFELQYISGRYCDIDINECTMDSHTCSHHANCFTQ 240
 QY 241 GSFKCKCKQYKGNGLRCSAIPENSVKVEVLAAPGTIKRIKLLAHKNSMKKAKIKVNT 300
 DB 241 GSFKCKCKQYKGNGLRCSAIPENSVKVEVLAAPGTIKRIKLLAHKNSMKKAKIKVNT 300
 QY 301 PEPTPTPKVNLDPFNVEEIVSRGNSHGKKGNEK 338
 DB 301 PEPTPTPKVNLDPFNVEEIVSRGNSHGKKGNEK 338

XX AAB18669;
 AC 22-JAN-2001 (first entry)
 XX
 DT
 XX Amino acid sequence of a human PRO320 polypeptide.
 DE
 XX Fibulin homologue; PRO320; PRO938; PRO1031; PRO296; PRO213; PRO1330;
 KW PRO1449; angiogenesis; cardiovascularisation; cardiovascular disorder;
 KW endothelial disorder; angiogenic disorder; cancer; trauma; wound;
 KW arteriosclerosis; cardiac hypertrophy.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key
 PI Peptide
 FT Location/Qualifiers
 FT 1..21
 FT /note= "signal sequence"
 FT 18..24
 FT /note= "N-myristoylation site"
 FT 21..27
 FT /note= "N-myristoylation site"
 FT 30..36
 FT /note= "N-myristoylation site"
 FT 44..50
 FT /note= "N-myristoylation site"
 FT 54..58
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 59..65
 FT /note= "N-myristoylation site"
 FT 68..74
 FT /note= "N-myristoylation site"
 FT 80..92
 FT /note= "EGF-like domain cysteine pattern signatu"
 FT 109..121
 FT /note= "aspartic acid and asparagine hydroxylation site"
 FT 114..120
 FT /note= "N-myristoylation site"
 FT 191..203
 FT /note= "aspartic acid and asparagine hydroxylation site"
 FT 196..202
 FT /note= "N-myristoylation site"
 FT 236..248
 FT /note= "aspartic acid and asparagine hydroxylation site"
 FT 241..247
 FT /note= "N-myristoylation site"
 FT 255..261
 FT /note= "N-myristoylation site"
 FT 326..332
 FT /note= "N-myristoylation site"
 FT 330..336
 FT /note= "N-myristoylation site"
 FT 330..334
 FT /note= "amidation site"
 XX
 XX W0200053752-A2.
 PN
 XX
 PD 14-SEP-2000.
 PD
 PE 30-DEC-1999; 99WO-US31274.
 PE
 XX
 XX 08-MAR-1999; 99WO-US05028.
 PR 21-APR-1999; 99US-0130232.
 PR 26-APR-1999; 99US-0131022.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 02-DEC-1999; 99WO-US28565.
 XX
 XX (GETH) GENENTECH INC.
 PA Baker KP, Ferrara N, Goddard A, Gurney AL, Hillan KJ, Williams PM;
 PI Wood WI;
 XX

DR WPI: 2000-638138/61.
 DR N-PSDB; AAAY5686.
 XX
 PT A composition useful for treatment and diagnosis of a cardiovascular,
 PT endothelial or angiogenic disorder, especially cancer, comprises (an
 PT agonist or antagonist of) a PRO320, PRO938, PRO1031, PRO296, PRO213,
 PT PRO1330 or PRO1449 polypeptide -
 XX
 PS Claim 67; Fig 2; 152pp; English.
 XX
 CC The present sequence represents PRO320, a fibulin homologue. The
 CC specification describes PRO320, PRO938, PRO1031, PRO296, PRO213,
 CC PRO1330 and PRO1449 polypeptides. The polypeptides promote or
 CC inhibit angiogenesis and cardiovascularisation in mammals. The
 CC polypeptides are used for the treatment and diagnosis of a
 CC cardiovascular, endothelial or angiogenic disorder, especially
 CC cancer. Disorders that can be diagnosed, treated or prevented by
 CC the polypeptides of the invention include trauma such as wounds,
 CC arteriosclerosis, and cardiac hypertrophy.
 CC
 XX
 SQ Sequence 338 AA;
 Query Match 63.1%; Score 1931; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPMSIALPLLSSWVAGFGNNAASARRHGLLASARQPGVCHYGTKLACCGWRNRNSKV 60
 DB 1 MPLPMSIALPLLSSWVAGFGNNAASARRHGLLASARQPGVCHYGTKLACCGWRNRNSKV 60
 QY 61 CEATCEGCKFGECVGNKRCRPFYGTGKTCSDVNECGMKRPRCOHRCVTHGSYKFC 120
 DB 61 CEATCEGCKFGECVGNKRCRPFYGTGKTCSDVNECGMKRPRCOHRCVTHGSYKFC 120
 QY 121 LSGHMLMPDATCVNSRRCAMINCOYSEDETEEGPQCLCPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRRCAMINCOYSEDETEEGPQCLCPSSGLRLAPNGRCLDIDECAS 180
 QY 181 GKVCPCPNRRCVNTFGSYCCCHIGFELYISGRYDCIDIECHMDSTGSHHANCFTWQ 240
 DB 181 GKVCPCPNRRCVNTFGSYCCCHIGFELYISGRYDCIDIECHMDSTGSHHANCFTWQ 240
 QY 241 GSFCKCKGKGYKNGLCRSALPENSVEVLRAPTIKDIRIKLLAHKNSMKKAKIKNVT 300
 DB 241 GSFCKCKGKGYKNGLCRSALPENSVEVLRAPTIKDIRIKLLAHKNSMKKAKIKNVT 300
 QY 301 PEPTRTPTPKVNLQPFNVEEIVSRGNSHGKKGNEEK 338
 DB 301 PEPTRTPTPKVNLQPFNVEEIVSRGNSHGKKGNEEK 338

RESULT 11
 AAAY5339
 ID AAAY5339 standard; Protein; 338 AA.
 AC AAAY5339;
 XX
 DT 25-SEP-2000 (first entry)
 XX
 DE Human PRO320 antitumour protein.
 XX
 KW PRO320; human; antitumour; tumour; therapy; cytostatic;
 KW breast cancer; ovarian cancer; renal cancer; colorectal cancer;
 KW uterine cancer; prostate cancer; lung cancer; bladder cancer;
 KW central nervous system cancer; melanoma; leukemia; neoplasm.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT Peptide 1..21
 FT Protein /label= Signal_peptide
 FT 22..338
 FT /label= PRO320

FT Modified-site 330..334
 FT /note= "amidation site"
 FT
 FT Modified-site 109..121
 FT /note= "Asn hydroxylation site"
 FT
 FT Modified-site 191..203
 FT /note= "Asn hydroxylation site"
 FT
 FT Modified-site 236..248
 FT /note= "Asn hydroxylation site"
 FT
 FT Region 80..91
 FT /note= "epidermal growth factor-like domain
 FT cysteine pattern signature"
 FT
 FT Domain 103..125
 FT /note= "calcium-binding epidermal growth
 FT factor-like domain"
 FT
 FT Domain 230..252
 FT /note= "calcium-binding epidermal growth
 FT factor-like domain"
 FT
 FT Domain 185..207
 FT /note= "calcium-binding epidermal growth
 FT factor-like domain"
 FT
 XX
 PN W0200037638-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 02-DEC-1999; 99WO-US28565.
 XX
 PR 22-DEC-1998; 98US-0113296.
 PR 08-MAR-1999; 99WO-US05028.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145638.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Marsters SA;
 PI Napier MA, Pitti RM, Wood WL;
 PI
 DR WPI: 2000-442668/38.
 DR N-PSDB; AAA49718.
 XX
 PT Novel composition to inhibit neoplastic cell growth or for treating
 PT tumor in mammal comprises polypeptides PRO179, PRO207, PRO320, PRO219,
 PT PRO221, PRO224, PRO328, PRO301, PRO356, PRO359 or
 PT PRO866 -
 XX
 PS Claim 19; Fig 6; 172pp; English.
 XX
 CC The present sequence is that of human antitumour protein PRO320, as
 CC deduced from a foetal lung cDNA clone (see AAA49718). PRO320 has a
 CC mol.wt. of 37,143 and a pI of 8.92. A claimed method for inhibiting
 CC the growth of a tumour cell comprises exposing the tumor cell
 CC to PRO179, PRO207, PRO320, PRO219, PRO221, PRO224, PRO328, PRO301,
 CC PRO356, PRO359 or PRO866 (see AAAY5337-49), their
 CC agonists or chimeric polypeptides incorporating them. The tumour
 CC is especially a cancer selected from breast, ovarian, renal,
 CC colorectal, uterine, prostate, lung, bladder and central nervous
 CC system cancer, melanoma and leukemia. Methods for the recombinant
 CC expression of the antitumour proteins are also provided.
 CC
 XX
 SQ Sequence 338 AA;
 Query Match 63.1%; Score 1931; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1.9e-134;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPMSIALPLLSSWVAGFGNNAASARRHGLLASARQPGVCHYGTKLACCGWRNRNSKV 60
 DB 1 MPLPMSIALPLLSSWVAGFGNNAASARRHGLLASARQPGVCHYGTKLACCGWRNRNSKV 60

QY 61 CEATCEPGCKFGECEVGPKNCRCPGYTGKTCSDVNECGMKPRPCQHRVCVNTHSGYKFC 120
Db 61 CEATCEPGCKFGECEVGPKNCRCPGYTGKTCSDVNECGMKPRPCQHRVCVNTHSGYKFC 120
QY 121 LSGHMLMPDATICVNSRTCAMINCQYSCDTEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
Db 121 LSGHMLMPDATICVNSRTCAMINCQYSCDTEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTMQ 240
Db 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTMQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNTV 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNTV 300
QY 301 PEPTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEK 338
Db 301 PEPTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEK 338

RESULT 12

AAM23677 standard; Protein: 331 AA.

AC AAM23677;
DT 12-OCT-2001 (first entry)

Human EST encoded protein SEQ ID NO: 1202.

Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
diagnostics; forensic test; gene mapping; genetic disorder;
biodiversity; gene therapy; nutrition.

Homo sapiens.

WO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02687.

25-JAN-2000; 2000US-0491404.

17-JUL-2000; 2000US-0617746.

03-AUG-2000; 2000US-0631451.

15-SEP-2000; 2000US-0663870.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

Cao Y, Drmanac RA, Zhang J, Werhman T;

WPI: 2001-476164/51.

N-PSDB: AAH98336.

isolated polypeptide for treatment of diseases, diagnostics, raising

antibodies and research use -

Claim 20; Page 868-869; 1275pp; English.

The present invention provides the protein and coding sequences of novel
proteins from a variety of organisms, including human, dog, cat, horse,
cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
urchin and tomato. These were derived from expressed sequence tags (ESTs)
from the organism of interest. They can be used in diagnostics,
forensics, gene mapping, identification of mutations, to assess
biodiversity and for nutritional purposes. The present sequence is a
protein of the invention.

Sequence 331 AA;

Query Match 61.9%; Score 1894; DB 22; Length 331;
Best Local Similarity 100.0%; Pred. No. 9,7e-132;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLPWSLALPLLISVAVGFGNANASARRHGLASARQPGVCHYGTKLACCYGWRNRSKV 60
Db 1 MPLPWSLALPLLISVAVGFGNANASARRHGLASARQPGVCHYGTKLACCYGWRNRSKV 60
QY 61 CEATCEPGCKFGECEVGPKNCRCPGYTGKTCSDVNECGMKPRPCQHRVCVNTHSGYKFC 120
Db 61 CEATCEPGCKFGECEVGPKNCRCPGYTGKTCSDVNECGMKPRPCQHRVCVNTHSGYKFC 120
QY 121 LSGHMLMPDATICVNSRTCAMINCQYSCDTEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
Db 121 LSGHMLMPDATICVNSRTCAMINCQYSCDTEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTMQ 240
Db 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFTMQ 240
QY 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNTV 300
Db 241 GSFCKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKDIRIKLLAHKNSMKKAKIKNTV 300
QY 301 PEPTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEK 331
Db 301 PEPTPTPKVNLQPFNYEEIVSRGNSHGKKGNEEK 331

RESULT 13

AAB01429 standard; Protein: 333 AA.

AAB01429;

20-OCT-2000 (first entry)

Murine TANGO 212.

TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;
graft versus-host disease; rheumatoid arthritis; psoriasis;
inflammatory bowel disease; septic shock; ulcerative colitis;
Crohn's disease; chronic myelogenous leukemia; cancer; liver
disease; Hodgkin's disease; osteoarthritis; Lyme's disease;
cachexia; autoimmune disease; myasthenia gravis; autoimmune
systemic lupus erythematosus; transgenic animal; diagnosis;
prognosis; prophylactic; therapeutic; mouse.

Mus musculus.

WO200039284-A1.

06-JUL-2000.

23-DEC-1999; 99WO-US31025.

30-DEC-1998; 98US-0223546.

(MTLL-) MILLENNIUM PHARM INC.

Holtzman DA;

WPI: 2000-465743/40.

N-PSDB: AAA47480.

Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,
224 and 239 polypeptides useful for the treatment of asthma, rheumatoid
arthritis, psoriasis and autoimmune diseases

Claim 8; Fig 28; 209pp; English.

Nucleic acids encoding TANGO polypeptides are useful as modulating

CC agents for regulating cellular processes like asthma, graft
 CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory
 CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,
 CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic
 CC lupus erythematosus. The nucleic acids are also useful for producing
 CC transgenic animals and the RANGO polypeptides themselves. Partial
 CC RANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in
 CC forensic biology, for diagnostic assays, prognostic assays,
 CC pharmacogenomics and for monitoring clinical trials. RANGO
 CC polypeptides are suitable for both prophylactic and therapeutic
 CC methods for treating a subject at risk of a disorder or having a
 CC disorder associated with aberrant RANGO expression. A wide range
 CC of cellular disorders can be treated.

XX Sequence 333 AA;

SO Query Match 48.7%; Score 1491; DB 21; Length 333;
 Best Local Similarity 76.2%; Pred. No. 5.2e-102;
 Matches 260; Conservative 27; Mismatches 46; Indels 8; Gaps 3;

QY 1 MRLPMSLALPRLLSVAVAGFGNMAASARHHGLASAROPGVCHYGTKLACCTGMRNSKGV 60
 Db 1 MOPWGLALPRLLPVAVAGVG--TSPWDYGLSALAHOGVCHYGTAKACCYGMKRNKNGV 58

QY 61 CEATCEPCRCFGECEGVPMKRCRCFPGYTGTCTCSQDVNECGMKPRPCQHRVCVNTHGSYKFC 120
 Db 59 CEAMCEPCRCFGECEGVPMKRCRCFPGYTGTCTCTODVNECGVAFRQCQHCNVTHGSYKFC 118

QY 121 LSGHMLPDAVCNSRTCAAMINCOYSCEDTEBGPCCLPSSGLRLAPNGRCLDIDECAS 180
 Db 119 LSGHMLPDAVCNSRTCAAMINCOYSCEDTEBGPCCLPSSGLRLAPNGRCLDIDECAS 178

QY 181 GKVICPYNRCVNTFGSYCKHIGFELYISGRYDCIDINECTDMSHTCSHHANCFMTQ 240
 Db 179 SKAVCPSNRCVNTFGSYCKHIGFELYISGRYDCIDINECALNTHPCSPHANCLMTR 238

QY 241 GSFCKCKOGYKGNLRCSAIPENSVEKLEAPGTIKDIRKLLAHKMSKKAKIKNVT 300
 Db 239 GSFCKCKOGYKGNLRCQSVIPEHSVKELIAPGTIKDIRKLLAHKTKMKKVKALNVT 298

QY 301 PEPTRTPEPKYNLOPNTVEEIVSRGNSHGKKGNEEKAKE 341
 Db 299 PRPASTRPVKYNL--PYSESEGVSRGRNYDG-----EQKKKK 333

RESULT 14
 AAY76110
 ID AAY76110 standard; Protein: 284 AA.

XX AC AAY76110;
 XX DT 27-MAR-2000 (first entry)
 XX DE Rat TGF-beta homologue, SEQ ID NO:389.
 XX KW Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;
 KW embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
 KW secreted; transmembrane; inflammation; cancer; neurological disease;
 KW angiogenesis; tumour vascularisation; growth disorder;
 KW developmental disorder; skin wound; hair follicle disorder;
 KW anti-inflammatory; cytostatic; neuroprotective; vulnery.

XX OS Rattus sp.
 XX PN MO9955865-A1.
 XX PD 04-NOV-1999.
 XX PF 29-APR-1999; 99WO-NZ00051.
 XX PR 29-APR-1998; 98US-0069726.

PR 09-NOV-1998; 98US-0188930.
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA Strachan L, Sleeman M, Watson JD, Onrust R, Kumbje A, Murison JG;
 XX WPI: 2000-072177/06.
 DR N-PSDB; AAZ61818.
 XX Novel polynucleotides useful for the treatment of various conditions
 PT including wounds and cancer -
 PS Claim 4; Page 220-221; 235pp; English.

CC The invention relates to novel nucleic acid sequences derived from rat
 CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
 CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying
 CC cells. Polypeptides of the invention may be used to treat inflammation,
 CC cancer and neurological diseases. The proteins may be used to stimulate
 CC the growth and motility of keratinocytes, to inhibit the growth of
 CC cancer cells, to modulate angiogenesis and tumour vascularisation, to
 CC modulate skin inflammation, to modulate epithelial cell growth and to
 CC inhibit binding of HIV-1 to leukocytes. The invention may also be used
 CC to treat growth and developmental defects, skin wounds and hair follicle
 CC disorders. Sequences AAY75942-Y76123 represent polypeptides encoded
 CC by cDNA sequences derived from several mouse, rat or human skin cell
 CC types. Sequences AAY75942-Y75947, AAY76020-Y76021, AAY76094-Y76104 and
 CC AAY76119 are proteins with an N-terminal signal sequence, indicating
 CC that they are secreted. Sequences AAY75986-Y75989, AAY76061-Y76071,
 CC AAY76106-Y76109 and AAY76121-Y76122 are proteins with one or more
 CC putative transmembrane domains.

XX Sequence 284 AA;

SO Query Match 42.1%; Score 1288; DB 21; Length 284;
 Best Local Similarity 82.6%; Pred. No. 4e-87;
 Matches 218; Conservative 14; Mismatches 30; Indels 2; Gaps 1;

QY 1 MRLPMSLALPRLLSVAVAGFGNMAASARHHGLASAROPGVCHYGTKLACCTGMRNSKGV 60
 Db 23 MOPWGLALPRLLPVAVAGVG--TSPWDYGLSALAHOGVCHYGTAKACCYGMKRNKNGV 80

QY 61 CEATCEPCRCFGECEGVPMKRCRCFPGYTGTCTCSQDVNECGMKPRPCQHRVCVNTHGSYKFC 120
 Db 81 CEAMCEPCRCFGECEGVPMKRCRCFPGYTGTCTCTODVNECGVAFRQCQHCNVTHGSYKFC 140

QY 121 LSGHMLPDAVCNSRTCAAMINCOYSCEDTEBGPCCLPSSGLRLAPNGRCLDIDECAS 180
 Db 141 LSGHMLPDAVCNSRTCAAMINCOYSCEDTEBGPCCLPSSGLRLAPNGRCLDIDECAS 200

QY 181 GKVICPYNRCVNTFGSYCKHIGFELYISGRYDCIDINECTDMSHTCSHHANCFMTQ 240
 Db 201 SKAVCPSNRCVNTFGSYCKHIGFELYISGRYDCIDINECTNTHPCSPHANCLMTR 260

QY 241 GSFCKCKOGYKGNLRCSAIPENSVEKLEAPGTIKDIRKLLAHKMSKKAKIKNVT 300
 Db 261 GSFCKCKOGYKGNLRCQSVIPEHSVKELIAPGTIKDIRKLLAHKTKMKKVKALNVT 284

RESULT 15
 AAB56049
 ID AAB56049 standard; Protein: 284 AA.

XX AC AAB56049;
 XX DT 08-MAR-2001 (first entry)
 XX DE Skin cell protein, SEQ ID NO: 389.
 XX KW Rat; skin cell; cytostatic; anti-inflammatory; anti-HIV;
 KW neurologic; neuroprotective; vulnery; immunomodulatory; vaccine;
 KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;
 KW inflammation; neurological disease.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 15:59:49 ; Search time 15 Seconds
(without alignments)
900,490 Million cell updates/sec

Title: US-09-687-860-24
Perfect score: 3060
Sequence: 1 MPELPSLALPLLSWVAGF.....VDGVLVSGICPDSLVSVD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	354.5	11.6	1833	US-08-479-722B-2	Sequence 2, Appli
2	354.5	11.6	1833	PCT-US95-02251-18	Sequence 18, Appli
3	339.5	11.1	956	US-08-897-443-3	Sequence 3, Appli
4	335.5	11.0	2703	US-08-185-432-19	Sequence 19, Appli
5	332.5	10.5	886	US-09-110-116-3	Sequence 3, Appli
6	332.5	10.5	2556	US-08-083-590A-20	Sequence 20, Appli
7	322.5	10.5	2556	US-08-532-384-20	Sequence 20, Appli
8	320.5	10.5	638	US-08-897-443-1	Sequence 1, Appli
9	320.5	10.5	2556	US-08-185-432-17	Sequence 17, Appli
10	320.5	10.5	1394	5177197-30	Patent No. 5177197
11	313	10.2	443	US-08-833-963C-2	Sequence 2, Appli
12	313	10.2	443	US-08-980-514-1	Sequence 1, Appli
13	313	10.2	448	US-08-884-072-1	Sequence 1, Appli
14	313	10.2	448	US-09-212-168-1	Sequence 1, Appli
15	311.5	10.2	2523	US-08-185-432-18	Sequence 18, Appli
16	310.5	10.1	810	US-08-820-170A-34	Sequence 34, Appli
17	310.5	10.1	810	US-09-055-699-34	Sequence 34, Appli
18	310.5	10.1	810	US-09-273-565-34	Sequence 34, Appli
19	310.5	10.1	810	US-09-565-538-34	Sequence 34, Appli
20	297.5	9.7	1218	US-09-214-278-7	Sequence 7, Appli
21	296.5	9.7	1010	US-08-882-046-7	Sequence 7, Appli
22	296.5	9.7	1036	US-09-068-740A-6	Sequence 6, Appli
23	296.5	9.7	1187	US-09-068-740A-7	Sequence 7, Appli
24	296.5	9.7	1218	US-08-400-159-6	Sequence 6, Appli
25	296.5	9.7	1218	US-08-611-729A-6	Sequence 6, Appli
26	296.5	9.7	1218	US-08-882-046-2	Sequence 2, Appli
27	296.5	9.7	1218	US-09-068-740A-11	Sequence 11, Appli

28	296.5	9.7	1219	US-08-882-046-5	Sequence 5, Appli
29	296	9.7	2471	US-08-185-432-16	Sequence 16, Appli
30	296	9.7	2471	US-08-083-590A-19	Sequence 19, Appli
31	296	9.7	2471	US-08-532-384-19	Sequence 19, Appli
32	293	9.6	337	US-09-188-930-186	Sequence 186, App
33	292	9.5	713	US-08-872-855-5	Sequence 5, Appli
34	291	9.5	387	US-08-884-072-5	Sequence 5, Appli
35	291	9.5	387	US-08-833-963C-9	Sequence 9, Appli
36	291	9.5	387	US-08-980-514-3	Sequence 3, Appli
37	291	9.5	387	US-09-212-168-5	Sequence 5, Appli
38	291	9.5	652	US-08-751-305-2	Sequence 2, Appli
39	288.5	9.4	1193	US-08-400-159-10	Sequence 10, Appli
40	288.5	9.4	1193	US-08-611-729A-10	Sequence 10, Appli
41	287.5	9.4	722	US-08-872-855-4	Sequence 4, Appli
42	287.5	9.4	722	US-08-981-392-12	Sequence 12, Appli
43	284.5	9.3	1248	US-08-882-046-6	Sequence 6, Appli
44	282	9.2	1065	US-08-400-159-8	Sequence 8, Appli
45	282	9.2	1212	US-09-214-278-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-479-722B-2
Sequence 2, Application US/08479722B

Patent No. 6074840

GENERAL INFORMATION:

APPLICANT: Bonadio, Jeffrey

APPLICANT: Yin, Mushan

TITLE OF INVENTION: LATENT TGF (BINDING PROTEIN (LTBP)

	Query Match	11.6%	Score 354.5	DB 3	Length 1833	
	Best Local Similarity	33.6%	Pred. No. 3e-20			
	Matches	99	Conservative	27	Mismatches	92
					Indels	77
					Gaps	19
QY	37	QPGVCHYT-----	KLACCYGRNRNSKGYCAETC---	PG-CKFGECV---	GPNKC	80
Db	947	QPGVSGGRGCSNTEGSSYNCECDRGIYMRKGCODINCEHRHGTGTPDGCNVCSPSGSYTCL				1006
QY	81	RCFPYATKTC-OPVNECGMKPRPCQH-RCVNTHGSYKFCFLSGHMLMPD-----				129
Db	1007	ACEEGYVGSGSCVDVNEC-LTPGCTHGRCLNMGSFRCSCPEGYEYTPDKKGGADYDE				1065
QY	130	-----AICVN-----SRTCAMINQY-----	SCEDTEGPGC-----	LCPS-----		160
Db	1066	CASRASCTGLCLNTEGSEFTCSACOSGAVWNEDGTACDELD---	CAPPGVCPGTVCINT			1122
QY	161	-----SGRLAPNGRDCLDIDECAAGKVICPYNNRCVNTFGSYCKRHIFETLY				210
Db	1123	VGSFSCKCDGGRPNPAGNRCDEVDCEGPOSSC-RGEGCKNTGSSYQCLDHQGFQL--				1179
QY	211	ISGRVDCIDINICTMDSHTSCSHNACFMTOGSGFRCKCKQGYGK--NGRCSAIP				263
Db	1180	VNGTM-CEVDNCEVEEH-CAPHGECILNSLGSFCLCAGPFAISAGGTRCQVDYE				1232

```

1      RESULT 2
2      PCT-US95-02251-18
3      ; Sequence 18, Application PC/TUS9502251
4      ; GENERAL INFORMATION:
5      ; APPLICANT:
6      ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
7      ; TITLE OF INVENTION: CELLS
8      ; NUMBER OF SEQUENCES: 18
9      ; CORRESPONDENCE ADDRESS:
10     ; ADDRESSEE: Arnold, White & Durkee
11     ; STREET: P. O. Box 4433
12     ; CITY: Houston
13     ; STATE: Texas
14     ; COUNTRY: United States of America
15     ; ZIP: 77210
16     ; COMPUTER READABLE FORM:
17     ; MEDIUM TYPE: Floppy disk
18     ; COMPUTER: IBM PC compatible
19     ; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
20     ; SOFTWARE: PatentIn Release #1.0, Version
21     ; SOFTWARE: #1.30
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: PCT/US95/02251
24     ; FILING DATE: CONCURRENTLY HEREMWITH
25     ; CLASSIFICATION:
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 08/316,650
28     ; FILING DATE: 30-SEP-1994
29     ; CLASSIFICATION:
30     ; APPLICATION NUMBER: US 08/199,760
31     ; FILING DATE: 18-FEB-1994
32     ; CLASSIFICATION:
33     ; ATTORNEY/AGENT INFORMATION:
34     ; NAME: Parker, David L.
35     ; REGISTRATION NUMBER: 32,165
36     ; REFERENCE/DOCKET NUMBER: UMIC009P-
37     ; TELECOMMUNICATION INFORMATION:
38     ; TELEPHONE: (512) 418-3000
39     ; TELEFAX: (713) 789-2679
40     ; TELEX: 79-0924
41     ; INFORMATION FOR SEQ ID NO: 18:
42     ; SEQUENCE CHARACTERISTICS:
43     ; LENGTH: 1833 amino acids
44     ; TYPE: amino acid
45     ; TOPOLOGY: linear
46     ; MOLECULE TYPE: protein
47     ; PCT-US95-02251-18

```

Query Match	11.6%: Score 354.5; DB 5; Length 1893;
Best Local Similarity	33.6%: Pred. NO. 3e-20;
Matches 99; Conservative	27; Mismatches 92; Indels 77; Gaps 19;
QY 37 QPVCVCHYT-----KLACCYGMRRNSKGYCEATCE--PG-CKFGEV--GPNKC-80	
Db 947 QPGVSGSGRCSNTEGSHYCEDRGYIMWRKGHODINCRHFGTCDPDGRCVSPSGSYTCL-1006	
QY 81 RCFPGYTKTCS-QDVNRCGKMKPRCOH-RCVNTGSKYKFCFLSGHMLMPD-----129	
Db 1007 ACEBEYVAGSGSCYDVNRC-LTPGCTHGRCLNMGSGSRSCSEPEEYVTPDKKGGRDVDE-1065	
QY 130 -----ATCVN---SRTCAMINQY-----SCEDTEEGPQC-----LCPS-----160	
Db 1066 CASRASCTGLCLNTEGSGFTCSACSGYVWEDGTACEDLE--CAEPGVCPPTVCNTT-1122	
QY 161 -----SGLRLAPAGRCGLDIDECASGAVICPYNRRVCNFTGSGYYCKCHIEFLQY-210	
Db 1123 VGSFSCKDDQDQGRPNPLGNRCEDVDECEBQSSC-RGEGCKNTEGSGYQCLCHQGFOL--1179	
QY 211 ISGRVCDIDINECTMSDHTSCSHNACFPMSGGSKKCKQGYKG--NGRCSAIRE-263	
Db 1180 VNGTM-CEDVNECGVEEH-CAPIEGCLNLSIGSFCLCAPGFSASAGRCQDVDE-1232	

```

1      RESULT 3
2      US-08-897-443-3
3      Sequence 3, Application US/08897443
4      Patent No. 5361263
5      GENERAL INFORMATION:
6      APPLICANT: Hillman, Jennifer L.
7      APPLICANT: Lal, Preeti
8      APPLICANT: Corley, Neil C.
9      APPLICANT: Shah, Purvi
10     APPLICANT: Kaser, Matthew
11     TITLE OF INVENTION: HUMAN MATRILIN-3
12     NUMBER OF SEQUENCES: 4
13     CORRESPONDENCE ADDRESS:
14     ADDRESSEE: Incyte Pharmaceuticals, Inc.
15     STREET: 3174 Porter Drive
16     CITY: Palo Alto
17     STATE: CA
18     COUNTRY: USA
19     ZIP: 94304
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Diskette
22     COMPUTER: IBM Compatible
23     OPERATING SYSTEM: DOS
24     SOFTWARE: FastSeq for Windows Version 2.0
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/897,443
27     FILING DATE: Filed Herewith
28     PRIOR APPLICATION DATA:
29     APPLICATION NUMBER:
30     FILING DATE:
31     ATTORNEY/AGENT INFORMATION:
32     NAME: Billings, Lucy J.
33     REGISTRATION NUMBER: 36,749
34     REFERENCE/DOCKET NUMBER: PF-0348 US
35     TELECOMMUNICATION INFORMATION:
36     TELEPHONE: 415-855-0555
37     TELEFAX: 415-845-4166
38     INFORMATION FOR SEQ. ID NO. 3:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 956 amino acids
41     TYPE: amino acid
42     STRANDEDNESS: single
43     TOPOLOGY: linear
44     IMMEDIATE SOURCE:
45     LIBRARY: GenBank
46     CLONE: 2072792
47     US-08-897-443-3

```

```

? RESULT 5
? US-09-110-116-3
? Sequence 3, Application US/09110116
? Patent No. 6013479
? GENERAL INFORMATION:
? APPLICANT: Xu, Hong
? APPLICANT: Cohen, Victoria L.
? APPLICANT: Stuart, Susan G.
? TITLE OF INVENTION: HUMAN ENK1-LIKE G PROTEIN COUPLED
? TITLE OF INVENTION: RECEPTOR
? FILE REFERENCE: PF-0550 US
? CURRENT APPLICATION NUMBER: US/09/110,116
? CURRENT FILING DATE: 1998-07-02
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 3
? LENGTH: 886
? TYPE: PRT
? ORGANISM: HOMO SAPIENS
?

```

```

; FEATURE:
; OTHER INFORMATION: 784994, GenBank
; US-09-110-116-3

Query Match      10.5%; Score 322.5; DB 3; Length 886;
Best Local Similarity 30.2%; Pred. No. 4.7e-18;
Matches 98; Conservative 29; Mismatches 112; Indels 85; Gaps 19;

QY 11 LLLSVAGGFGNMAARHGLASAROPV---CHYGTKLACGWMRNSKGVCEATCE 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 LLLFW-----GCCVWHSEGHIRPTKRNPKGNCRDST-LCPAVALCTNVTDSYTCCK 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 PG-----CK-FGEC-----VGPN-----KCCFCFGYTKT--- 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 QGLSSNGQNHKDPGVAKCIDECSSQPOPQSGNSSCKNLGSRKCKCLDGFSSPTGND 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 91 -----OSQVNEGCMKPRPC-QHR-CVNHGYSYKCFCLSGHMLMPLATCVNSRTC 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 WPGKPGNFSCPT-DINEC-LTSRVCPESHSDCVNSKSGYSCSQVG-FISRNSTCEDVNEC 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 139 AMINC---QYSCEDTEEGPQCLC---PSSGLRLAPNG--RDCLDIDECASGVICPYNR 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 ADPRACPRHATCNMTVGVYSCFCNPGFESSGHLSCGLKASCEIDECTE---MCPINS 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 RCVNTFSGYTKCHIGF-----ELQYISGRYDCIDINECTMDSHTSCHHANCENTQGSFK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 TCTMTPGSYFCTCHGFGFAPSSGGLNFTDQGYECRDIDECRODPSTCGPNSICTNALGSYS 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 CKCKGQYKNGLRCSAIPENSYKE 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 CGCTVGFHFN-----DEGSOKD 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
; US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown

```

```

; MOLECULE TYPE: peptide
; US-08-083-590A-20

Query Match      10.5%; Score 322.5; DB 1; Length 2556;
Best Local Similarity 25.3%; Pred. No. 1.9e-17;
Matches 94; Conservative 31; Mismatches 108; Indels 139; Gaps 16;

QY 36 RQPGYCHYT---KLACCYWMRNSKGVCE---ATCEPGCKEGEV---GPNKRCRFP 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 RNGTCDDLTLLEYKRCRCPGWSGKS---COADPCASNPCANGGCLPFEASYICHCP 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 85 GYTKTCSQDVNEGCMKRRPQH--RCVNTHGSYKCFELSGH----- 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 SFHGPTCRQDVNEGQKPRLCRHGGTCHNEVGSYKVCRAHTGPNCRPYVPSPCQ 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 -----MLMPAT-----CVNSRTCA-----MINC----- 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 NGGTCRPTGYDTHHCACLPFGTQNCENIDDCPGNNCKNGACVGYNTTNCPCPPERT 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 -QYSCEDTEE-----GPOCLPSSGLRLAPNGRDCLD-IDECASGK 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 GQYCTEDVDEQGLMPNACQNGGTCHNTHGYNVCVNGW-----TGDCSENIIDCAS-- 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 VICPNRRCVNTFSGYTKCHIGF-----FELQYISGRYDCI-- 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 AACFHGATCHRVASFYCECFHGRGLCHLNDACISNPKNEGSCNDNPNVKAICTCP 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 -----DINECTMDSHTSCHHANCENTQGSFKCKGQYKNGLRCSAIPENSYKE 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 SGYTPACSDVDECSLAPNCEHAGKCIINTLGSFECCLOGY--TGPRCEIDVNECVSN 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 269 VLRAPTIKDRI 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 PCQNDATCLDQI 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
; US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown

```

```

; INFORMATION FOR SEQ ID NO: 20:
;
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2556 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: unknown
;     MOLECULE TYPE: peptide
;
US-08-532-384-20

```

Query Match	10.5%	Score 322.5	DB 3	Length 2556
Best Local Similarity	25.3%	Pred. No. 1.9e-17		
Matches 94	Conservative 31	Mismatches 108	Indels 139	Gaps 16

```

QY      36  ROPVCHGHT-----KLACCYGMRRNSRGVE---ATCEPGCKKFGECV---GPNKCCFP  84
Db      112  RNGSTCDLLLTLETKCKCNPMSKS---CGQADPCASNCANCGCLLPREASTICHCP  168

QY      85  GYTKTSQDVNECCMKPRPCQH--KCVNTHGSIYKFCLSGH-----              124
Db      169  SFHGPTCRQDVNECGOKPRLCRHGSGTCHNEGVSYRCVCRATHHGPNCERPYPCSPCQ  228

QY      125  -----MLMPAT-----CVMSRTCA--*MNC-----              143
Db      229  NGGTCRPTGDVTHIEACLPGETTGNCCEENIDDCPGNNCKKGACVDPVNTNYNCPCEWT  288

QY      144  -QYSCDEPTE-----GPOCIPSSGLRLAPNGRDCID-IDECAASG  182
Db      289  GCYCTEDVDECOLMPNMACQNGGTCHTNHGGYNCYCVNGW---TGHDCESENIDDCAS--  341

QY      183  VICPYNRRCVNTFGSIYCKCHIG-----FELQYISGRYDCT--  218
Db      342  AACPHGATCHDRVASVFCOECPHRTGTLCHLDNACISNPNESNGCOTNPNVNAKALCTCP  402

QY      219  -----DINECTDSHTCSHIANCFNTQGSFCKCKCKQGYKGNGLRCSAIPENSYKE  268
Db      402  SGYTGPAACSQDVDEDCSIGANPCENHAGKICNTLGSFFECOCLOGY--TSPRCEIDVNECVSN  459

QY      269  VLRAGTKIDRI  280
Db      460  PCQNDATFCLDRI  471

```

RESULT 8
US-08-897-443-1
Sequence 1, Application US/08897443
Patent No. 5981263
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
APPLICANT: Kaser, Mathew
TITLE OF INVENTION: HUMAN MATRILIN-3
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,443
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

```

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0348 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN02
CLONE: 681719

```

Query Match	10.5%	Score 320.5	DB 2	Length 638
Best Local Similarity	34.3%	Pred. No. 4.4e-18		
Matches 80, Conservative	27	Mismatches 91	Indels 35	Gaps 14

[illegible]

RESULT 9
 US-08-185-432-17
 Sequence 17 Application US/08185432
 Patent No. 5750652
 GENERAL INFORMATION:
 APPLICANT: Artavanis-Tsakonas, Spyridon
 APPLICANT: Bussseau, Isabelle
 APPLICANT: Diederich, Robert J.
 APPLICANT: Xu, Tian
 APPLICANT: Matsuno, Kenji
 TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND
 TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/185,432
 FILING DATE: 21-JAN-1994
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-17

Query Match 10.5%; Score 320.5; DB 1; Length 2556;
Best Local Similarity 25.3%; Pred. No. 2.8e-17;
Matches 94; Conservative 31; Mismatches 108; Indels 139; Gaps 16;

36 ROPGVCHGT-----KIACCYGMRNRKSGVCE-----ATCEPCCKFEGCY---GPNKCRGCFP 84
112 RNGTCDLLTLTEYKCRPRPMGSGKS---CGADPCASNPANGGOCCLPFESYICHCHPP 168
85 GYTKGTSQDVNECGMKRPPCOH--RCVNTGSKYKCFCLSGH-----124
169 SFHGPTCQWQDVNECGQKRLCRHGCTCHNEVGYRCVCRATHTGPNCEMPYVPCSPSCQ 228
125 -----MLMPDAT-----CVNSRTCA---MINC-----143
229 NGGTCRPTGDTVTHCACLPGFTGQNCENIDCGENMCKNGACVDGVNTYNCPCPPPWT 288
144 -QYSCDETE-----GPOCLCPSSGLRLAPNGRDCLD-IDECASGK 182
289 GQYCTEDVEDQCLMPNACQNGTCHNTHGYNCVCVNGM-----TGEDCSEHIDDCAS-- 341
183 VICYNRCVNTFGSYKCHIG-----FELQYISGRIDCI-- 218
342 AACFHGATCHDRVASFYCECPHGRTGLCHLDACISNPNCEGNSCDTNPVNGKAITCPC 401
219 -----DINECTMDSHNCFNTQGSFCKCKQGYKNGGLRCSAIPENSYKE 268
402 SGYTPRACSDQVDECSLGCANPCENHAGKCNITLGSFECCOLOGY--TGPRCELDVNECVSN 459
269 VLRAPGITKRI 280
460 PCQNDATCLDQI 471

RESULT 10
517197-30
Patent No. 517197
APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
WENSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
LENA; HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/487,343
FILING DATE: 27-FEB-1990
SEQ ID NO: 30
LENGTH: 1394
517197-30

Query Match 10.5%; Score 320; DB 6; Length 1394;
Best Local Similarity 31.2%; Pred. No. 1.4e-17;
Matches 89; Conservative 23; Mismatches 93; Indels 80; Gaps 16;

41 CHYGTKLACGYMRNRKSGVCE-----ATC-----EPG-----CRFG-----72
654 CEY-----CDSGYRMTQGRCEIDECINPSTCPDEQCVNPGSYQVCPTEGFGWNGQ 708
73 -----ECYGRN-----KCRCPFGYT-----GKTCSDVYNECGMKRPPCQHRC 109
709 CLADVDECLPVPYCANCDGNSLEGSYMSCHKGYRTPTDHKHC-RDIDECQOQNLGVNGQC 767
110 VNTGSKYKCFCLSGHMLMPDA-----TCVNSRTCAMINQYSCDETEEGPOCLCPSSG 162

Db 768 KNTGSEFRCTCGGQGYOLSAKDOCEDIDECOHRLICA-----HGOCNTEGSPQVC--DQG 822
Qy 163 LRIAPNGRCDIDIDECASGKVIQPNRRCVMTGPGSYKCHIGFELQYISGRIDCIDE 222
Db 823 YRASGLDGHCEIDINECLDKSYC-ORGCITNAGSYDCTCTDGFOL---DNKTCQDINE 878
Qy 223 CTMDSH--TCSHANCFNTQGSFCKCKQGY--KNGGLRCSAIP 263
Db 879 C---EHPGLCGPQGECLNTEGSHFHCVCQGFISADGRTCEIDIDE 920

RESULT 11
US-08-833-963C-2
Sequence 2, Application US/08833963C
Patent No. 5916769

GENERAL INFORMATION:
APPLICANT: Olsen, et al.
TITLE OF INVENTION: Extracellular/Epidermal Growth Factor
NUMBER OF INVENTION: HCA8A58x
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Ave
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,963C
FILING DATE: 11-Apr-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/05033
FILING DATE: 10-Apr-1996
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PP258
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8439
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 443 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-833-963C-2

Query Match 10.2%; Score 313; DB 2; Length 443;
Best Local Similarity 35.8%; Pred. No. 1.1e-17;
Matches 69; Conservative 28; Mismatches 78; Indels 18; Gaps 10;

Qy 77 PNKCRCPFGY--TGKTSQDVNECGMKRPPCO--HRCVNTGSKYKCFCLSGH-MLMPDAT 131
Db 106 PNPCH--PGYERDDDDSCVDVDECAQALHDCRPSDDCHNLGSGYQCTCPDQYRKIGPE-- 161
Qy 132 CVNSRTCAMINQYSCDETEEGPOCLCPSSGLRLAPNGRCDIDIDECASGKVIQPNRRC 191
Db 162 CVDIDECRYRCQHGRCVNLPGSFRCQC-EPGFOLGPNNRSCVDVNECDMG--APCEQRC 217
Qy 192 VNTGSKYKCHIGFELQYISGRYDCIDINECTMDSHNCFNTQGSFCKCKQGY 251
Db 218 FNSGTFELCRHQGYELH--RDGFSCLDIDECSSYSLQY--RCVNEBGRFSCHPQGY 273
Qy 252 KNGGLR-CSAIP 263
Db 274 QLATRLCQDIDE 286


```

; Patent No. 6303765
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTED for windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/212,168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/884,072
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0333 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 448 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: CORNNOT01
; CLONE: 45517
; US-09-212-168-1

Query Match      10.2%; Score 313; DB 4; Length 448;
Best Local Similarity 30.6%; Pred. No. 1,1e-17;
Matches 86; Conservative 36; Mismatches 127; Indels 32; Gaps 13;

QY 7 LALFLLSWVAGFGGNASAHHGLASAROP-GVCHTGT---KLACCYGRNRKSKGYCE 62
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 67 LCIFRTNPVYRGPSYNPSTYSGPYAPPLSAPNPPTISRPILICFQYOMDESNOCV 126
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 63 ATCEPCKCFEGC-----VGPNCRCFPGY---TGKTCSDYNECGMKPRPCQHRVCY 110
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 DYDECAIDSHOCNPTQICINTEGCTSCDTGWLLEQCL--DIDEC--RYGTCQOLCA 182
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 111 NTHGSYKCFCLSGHMLMPDA--TCVNSRTCAMIN--QYSCEDTEREGPOCLCPSSGLRLAPN 168
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 183 NVPSYSYCTCNPFTLNEDGRSCQDVNDECATENPCVQCVNTYSFICRC--DPGYELEED 241
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 169 GRCGLDIDECASGVICPYNRCVNTFGSYKCHIGFELQYISGRYCIDINECTMDSH 228
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 242 GVNCHSDMECESEFLQOH--ECVNOPTQYFCSCPFGYIL--LDDNRSQODINCECHRNH 297
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 229 TCSHHANCFNTGGSFKC-----KCKQGY-KGNGLRCSAIPEN 264
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 298 TCNLOQTCTNLOGGFKCIDPIRCEPILRISDNRCMCPAEN 338
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
US-08-185-432-18
; Sequence 18, Application US/08185432
```

```

; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Bussseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2523 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-185-432-18

Query Match      10.2%; Score 311.5; DB 1; Length 2523;
Best Local Similarity 29.8%; Pred. No. 1.5e-16;
Matches 89; Conservative 26; Mismatches 103; Indels 81; Gaps 18;

QY 39 GVCHY---GTRKACCYGM-----RRNSKGYCEATCEPG-----CKFGE----- 73
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 309 GTCHNTYGYNCYCVNMTGSEDCSENIDDCANAAHSGATCHDRAVASTYCCPHGRIGLL 368
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 74 -----CV-----GPNKRCFPGYTGKTCSDYNECGMKPRPCQHR--RC 109
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 369 CHLDNACISNFCNBSNCDTNPVNGKAICTCPGTYGPACNNVDYDECSLGANPEHGGRC 428
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 110 VYTHGSYKCFCLSGHMLMPDTC--VNSRTCAMINCO--YSEDTIEBGPQCLC-PS-SGLR 164
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 429 TMTLGSFQCNCPOGY---AGPRCEIDVNECLSNPCQNDSTGLDIDIGEFQCLCMGYEGLY 485
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 165 LAPNGRDLIDIDECASGVICPYNRCVNTFGSYKCHIGFELQYISGRYCIDINECT 224
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 486 CETN-----IDECASNP--CLHNGKCIDKINERCDPPTF-----SGNLQCHDPDECT 532
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 225 MDSHTCSHHANCFNTGGSFKCKQGYKGNGLRCSAIPENSVKRYLAP--GTIKDRI 280
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 533 --STPKNGAKCLDGPNSYTCQCTEGFTGR--HC-----EODINCEIPDPCHYGTCKDGI 583
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Search completed: October 17, 2002, 16:02:15
Job time : 23 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 15:57:58 : Search time 22 Seconds
(without alignments)
2415.336 Million cell updates/sec

Title: US-09-687-860-24
Perfect score: 3060
Sequence: 1 MFLPMSIALPLLISWVAGF.....VDGVLLVSGLCPPDLSLVDD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	3016	98.6	558	2 T17324	hypothetical prote
2	410.5	13.4	2907	2 A57278	fibritillin-2 precu
3	409.5	13.4	2918	2 A54105	fibritillin-2 precu
4	387.5	12.7	2871	2 A55567	fibritillin-1 - bovi
5	385.5	12.6	2871	2 A55567	fibritillin-1 precu
6	383.5	12.5	3002	2 A47221	fibritillin-1 precu
7	382	12.5	1221	2 A49457	fibritillin-2 precu
8	365.5	11.9	1574	2 T13954	MGF6 protein - ra
9	362.5	11.8	1184	2 A55184	fibritillin-2 precu
10	353.5	11.6	1620	2 T27283	hypothetical prote
11	350	11.4	683	2 C36346	fibritillin 1 precu
12	348	11.4	685	2 S78040	fibritillin 1 splice
13	338	11.0	601	2 B36346	fibritillin 1 precu
14	335.5	11.0	2703	2 A24420	notch protein - fr
15	335	10.9	705	2 S34968	fibritillin, splice
16	332	10.8	2437	2 S42612	transmembrane prot
17	331	10.8	589	2 T43210	fibritillin-ID precu
18	331	10.8	689	2 T42760	fibritillin, splice
19	330	10.8	712	2 T42990	fibritillin 1, splice
20	330	10.8	798	2 T22793	hypothetical prote
21	329	10.8	1712	2 A38261	masking protein
22	327.5	10.7	2318	2 S45306	notch3 protein - h
23	327.5	10.7	2321	2 S78549	notch3 protein - h
24	326	10.7	741	2 T46488	hypothetical prote
25	326	10.7	3507	2 T34513	hypothetical prote
26	325.5	10.6	1820	2 A54494	latent transformin
27	325	10.6	2531	2 S18188	notch protein homo
28	322.5	10.5	886	2 A57172	probable hormone r
29	322.5	10.5	2555	2 A40043	notch protein homo

30	322	10.5	2531	2 A46019	Notch-1 protein -
31	320	10.5	1394	2 A35626	transforming growt
32	315	10.3	1964	2 T09059	notch4 - mouse
33	311	10.2	810	2 T10756	Nel-homolog protei
34	307.5	10.0	2471	2 A49128	cell-fate determin
35	303.5	9.9	2524	2 A35844	notch protein - Af
36	296.5	9.7	1220	2 A56136	jagged protein pre
37	294.5	9.6	2139	2 A35672	crumbs protein - f
38	291.5	9.5	1081	2 T31329	receptor tyrosine
39	291	9.5	387	2 J38449	extracellular prot
40	290	9.5	493	2 JC5621	epidermal growth f
41	289.5	9.5	2352	2 T30201	Notch homolog prot
42	287.5	9.4	722	2 T48324	DETRA-like 1 - mou
43	280	9.2	1106	2 T18739	hypothetical prote
44	279	9.1	1251	2 A57293	latent transformin
45	278	9.1	3623	2 T09456	intrinsic factor-B

ALIGNMENTS

RESULT 1

T17324
hypothetical protein DKFZp564P2063.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17324
R:Ducrest-Hoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

A:Reference number: 218727
submitted to the Protein Sequence Database, September 1999

A:Accession: T17324

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-558 <DUE>

A:Cross-references: EMBL:AL117610

A:Experimental source: fetal brain; clone DKFZp564P2063

C:Genetics:

A>Note: DKFZp564P2063.1

Query Match	98.6%	Score 3016;	DB 2;	Length 558;
Best Local Similarity	98.9%	Pred. No. 2.8e-187;		
Matches 547;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
OY	1	MFLPMSIALPLLISWVAGFGNAAARHGLAASAROPGVCHYGTKLACCGMRNSGV	60	
DB	6	MFLPMSIALPLLISWVAGFGNAAARHGLAASAROPGVCHYGTKLACCGMRNSGV	65	
OY	61	CEATCEPCCKFGECEVGPNNKCRCPGYTGKTCSDVNECGMKPRPCQHRVCVTHGSKFCG	120	
DB	66	CEATCEPCCKFGECEVGPNNKCRCPGYTGKTCSDVNECGMKPRPCQHRVCVTHGSKFCG	125	
OY	121	LSGHHMLMDATCVNSRTCAAMINCOYSCEDTEBPGCLPSSGLRLAPNGRCCLDIDECAS	180	
DB	126	LSGHHMLMDATCVNSRTCAAMINCOYSCEDTEBPGCLPSSGLRLAPNGRCCLDIDECAS	185	
OY	181	GVKICPYNRCVNFPGSYGCKHIGFELYOTISGRDCCIDINECTMDSTSHHACFNTQ	240	
DB	186	GVKICPYNRCVNFPGSYGCKHIGFELYOTISGRDCCIDINECTMDSTSHHACFNTQ	245	
OY	241	GSFKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT	300	
DB	246	GSFKCKQGYKGNGLRCSAIPENSVEVLRAPGTIKRIKLLAHKNSMKKAKIKVNT	305	
OY	301	PEPTPTPPKYNLOPENTEEIYSRGNSHGKKGKNEEMKBSLEBEKREKALKNDIER	360	
DB	306	PEPTPTPPKYNLOPENTEEIYSRGNSHGKKGKNEEMKBSLEBEKREKALKNDIER	365	
OY	361	SLRGVFPFKVNEGEFGLIVORKALTSKLEHKDLNLSVPCSFHNGICDMKODREDDFD	420	
DB	366	SLRGVFPFKVNEGEFGLIVORKALTSKLEHKDLNLSVPCSFHNGICDMKODREDDFD	425	
OY	421	WNPADRDNAIGFYMAVPLAAGHKDGRILKLLPDLQPSNFCLLFDYRLAGDKVKLRV	480	
DB	426	WNPADRDNAIGFYMAVPLAAGHKDGRILKLLPDLQPSNFCLLFDYRLAGDKVKLRV	485	

Db 426 WNPDRDNAIFYAVAPVLACHKKDIDGRLKLLPDLQSQNSFCFLFYRLAGDKVKGLRV 485
 Oy 481 FVKNSSNNAIAEKTSTSEDEKMKTKIOLYOGTDATKSIIFEAERCKGTGELAVDGLLV 540
 |||||||
 Db 486 FVKNSSNNAIAEKTSTSEDEKMKTKIOLYOGTDATKSIIFEAERCKGTGELAVDGLLV 545
 Oy 541 SGLCPDLSLLSD 553
 |||||||
 Db 546 SGLCPDLSLLSD 558

RESULT 2

A57278
 fibrillin-2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 11-Jan-2000
 C:Accession: A57278
 R:Zhang, H.; Hu, W.; Ramirez, F.
 J. Cell Biol. 129, 1165-1176, 1995
 A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular
 A:Reference number: A57278; MUID:95263670
 A:Accession: A57278
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2907 <ZHA>
 A:Cross-references: GB:U39790; NID:9762830; PIDN:AAA74908.1; PID:9762831
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1239-1274/Domain: EGF homology <EGF>
 F:2488-2523/Domain: EGF homology <EGF>

Query Match 13.4%; Score 410.5; DB 2; Length 2907;
 Best Local Similarity 28.4%; Pred. No. 2.2e-18;
 Matches 95; Conservative 37; Mismatches 91; Indels 111; Gaps 13;

Oy 38 PGVCHYGT-----KLACCYGW-----RRNSKVCATCEPG-CRKEGCV---GPNKC 80
 |||||
 Db 1074 PGMCTYGCRTMTIGSFKCRNNGFALDMERNCTDIDECRISPDLGSGGICVNTGSEFC 1133
 |||||
 Oy 81 RCFPGY-----TGKTCQ-----D 94
 |||||
 Db 1134 ECFEYEGSFMMKMKMIDCECERNPLCRGTCTVNTGSPQCDPLGHELSPSREDCV 1193
 |||||
 Oy 95 VNECGMKRPPCOH-RCVNTGHSYKFCCLSGHMLMPDAT-CVNSRTCAMIN--CQSCEDT 150
 |||||
 Db 1194 INECSLSNLCRNKCVNMIGTYQCSCNPGYQATPDROGCTDIDECIMNNGCDTQCTNS 1253
 |||||
 Oy 151 EEGPQCLPSSGLRIAPNGRCLDIDECASGKVI-----PYNRRCV----- 192
 |||||
 Db 1254 EGSYECSC-SSEYALMPDGRSCADIDECENNPDCDGGQCTNIPETRYCLCYDGFMSMD 1312
 |||||
 Oy 193 -----NTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHT 229
 |||||
 Db 1313 MKTCDVNECDLNPINICFGECEENTKGSFICHQGLGYSVK--KGTGCTDVECEIGAHN 1370
 |||||
 Oy 230 CSHHANCFTNGSFCKCKGKGTGKGLRCSAIP 263
 |||||
 Db 1371 CDMHASCINLPISGFSKCRGEGWNGIKCIDLDE 1404

RESULT 3

A54105
 fibrillin-2 precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 17-Nov-2000
 C:Accession: A54105; S17063; S31101
 R:Zhang, H.; Apfelfroth, S.D.; Hu, W.; Davys, E.C.; Sanguinetti, C.; Bonadio, J.; Mecham, J.
 J. Cell Biol. 124, 855-863, 1994
 A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component prefer
 A:Reference number: A54105; MUID:94165150
 A:Accession: A54105
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2918 <ZHA>

A:Cross-references: GB:U03272
 R:Lee, B.; Godfrey, M.; Vitale, E.; Horii, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras, N.
 Nature 352, 330-334, 1991
 A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two diff
 A:Reference number: S17062; MUID:91304567
 A:Accession: S17063
 A:Molecule type: mRNA
 A:Residues: 752-1489, 1791, 'GS', 1794-1796, 'QLT', 1922-1923, 'LD', 1926, 'P', 1928 <LEE>
 A:Cross-references: EMBL:X62009
 R:Miliewicz, D.M.
 submitted to the EMBL Data Library, December 1992
 A:Reference number: S31101
 A:Accession: S31101
 A:Molecule type: mRNA
 A:Residues: 752-1407, 'R', 1409-1489, 1791, 'GS', 1794-1796, 'QLT', 1922-1923, 'LD', 1926, 'P',
 A:Cross-references: EMBL:X62009
 C:Genetics:

A:Gene: GDB:FBN2
 A:Cross-references: GDB:128122; OMIM:121050
 A:Map position: 5q23-5q31
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: extracellular protein
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-2918/Product: fibrillin-2 #status predicted <MAT>
 F:1245-1280/Domain: EGF homology <EGF>
 F:1970-2013/Domain: EGF homology <EGF>

Query Match 13.4%; Score 409.5; DB 2; Length 2918;
 Best Local Similarity 28.6%; Pred. No. 2.5e-18;
 Matches 96; Conservative 37; Mismatches 92; Indels 111; Gaps 13;

Oy 38 PGVCHYGT-----KLACCYGW-----RRNSKVCATCEPG-CRKEGCV---GPNKC 80
 |||||
 Db 1080 PGMCTYGCRTMTIGSFKCRNNGFALDMERNCTDIDECRISPDLGSGGICVNTGSEFC 1139
 |||||
 Oy 81 RCFPGY-----TGKTCQ-----D 94
 |||||
 Db 1140 ECFEYEGSFMMKMKMIDCECERNPLCRGTCTVNTGSPQCDPLGHELSPSREDCV 1199
 |||||
 Oy 95 VNECGMKRPPCOH-RCVNTGHSYKFCCLSGHMLMPDAT-CVNSRTCAMIN--CQSCEDT 150
 |||||
 Db 1200 INECSLSNLCRNKCVNMIGTYQCSCNPGYQATPDROGCTDIDECIMNNGCDTQCTNS 1259
 |||||
 Oy 151 EEGPQCLPSSGLRIAPNGRCLDIDECASGKVI-----PYNRRCV----- 192
 |||||
 Db 1260 EGSYECSC-SSEYALMPDGRSCADIDECENNPDCDGGQCTNIPETRYCLCYDGFMSMD 1318
 |||||
 Oy 193 -----NTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHT 229
 |||||
 Db 1319 MKTCDVNECDLNSINICFGECEENTKGSFICHQGLGYSVK--KGTGCTDVECEIGAHN 1376
 |||||
 Oy 230 CSHHANCFTNGSFCKCKGKGTGKGLRCSAIP 265
 |||||
 Db 1377 CDMHASCINLPISGFSKCRGEGWNGIKCIDLDECS 1412

RESULT 4

A53567
 fibrillin I - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
 C:Accession: A53567
 R:Titstra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
 Genomics 23, 460-485, 1994
 A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization
 A:Reference number: A53567; MUID:95137597
 A:Accession: A53567
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2871 <TIL>
 A:Cross-references: GB:U28748; NID:9508427; PIDN:AAA74122.1; PID:9508428
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1201-1236/Domain: EGF homology <EGF>

```

Query Match          12.7%; Score 387.5; DB 2; Length 2871;
Best Local Similarity 30.6%; Pred. No. 6.5e-17;
Matches 89; Conservative 34; Mismatches 91; Indels 77; Gaps 10;

QY      49 CCYGMRRSKVCYEATCEPG-----CKFGECV--GPNKCRC 82
       ||| | | | | | | | | | | | | | | | | | | | | |
Db      1081 CGRGCVMTPTDGFBECKDCDEGYESSGPMAMKNCMIDECQRDLPLCKGVCILNTESSYRBC 1140

QY      83 PPGY---TGKTCSODVNECGMKPRPCQH-RCVNTHGSKYCCFLSGHMLMPAT-CVNSRT 137
       ||| : | | | | | | | | | | | | | | | | | |
Db      1141 PPGHLPANISACIDINCELSAHLCPHGRCVNLIGKYQCACNPQYHSHPDLFCVDIDE 1200

QY      138 CAMIN-COYSEDTPEEBPOCLCPSSGLRLAPANGDCIDLDECASGAKYIC-----PY 187
       ||::: | | | | | | | | | | | | | | | | | |
Db      1201 CSIMNGGETCTCTNSEGSYESC-QPFGALMPDDQSCSTDICECENPNPICGGGACTNIIFG 1259

QY      188 NRRCY-----NTFSGYSXCKCHIGELIYISGR 214
       ||: | | | | | | | | | | | | | | | | |
Db      1260 EYKLCCTGYFMASEDMKTCVDVNECDLNPNICLSTGCENTGTGSPFICHCDMGISG--KGK 1317

QY      215 YDCIDINETMDSHFTCSHHANCFQTQSFFKCKCKGKYGKGNLCSAIPENS 265
       ||||| : : | | | | | | | | | | | | | | | |
Db      1318 TGCITDINCELGIAHNCDRHAVCTNAGSFKSCSGSMIGDGIKCTDIIDBES 1368

```

```

RESULT 5
A55624
Fibrillin-1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 11-Jan-2000
C:Accession: A55624
R:Yin, W.; Smiley, E.; Germiller, J.; Sangunetti, C.; Lawton, T.; Perelra, L.; Ramirez,
J. Biol. Chem. 270, 1798-1806, 1995
A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin gene
A:Accession: A55624
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <YIN>
A:Cross-references: GB:L29454; NID:g575509; PIDN:AAA56840.1; PID:g575510
C:Genetics:
A:Gene: Fbn-1
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:1201-1236/Domain: EGF homology <EGF>

Query Match 12.6% Score 385.5; DB 2; Length 2871;
Best Local Similarity 32.2%; Pred. No. 8.8e-17;
Matches 86; Conservative 34; Mismatches 62; Indels 65; Gaps 11;

QY 39 GVCHIGTKIACCGVMRRSKGVCVCEATCEPG-----CKRGECV 75
      |||
      |||
Db 1127 GICH-----NTEGYRCECPRHQLSPHLSNISACIDINCELSANLCPIGRCVN 1173
      |||
      |||

QY 76 --GPNKCKCFPEY---TAKTCSQDVNECGMKPRPCQHRVNTHGYSKPCFSLGMLMPDA 130
      |||
      |||
Db 1174 LIGKYOCAKPNPIYHPTHRFLFCVDIDECSIMNGCECFCTNSDGSFECSCOPGFALMPD 1232
      |||
      |||

QY 131 TCVNRRTCAMTINCOYSCDETE--EGPQ-----CLCPSSGRLAPNGDDCIDDEC 178
      |||
      |||
Db 1233 -----QRSTCID--QCEBPNPICDGGCTNIPGIRCLC-YDGFMASEDMMTCVDVNEC 1284
      |||
      |||

QY 179 ASGVICPEYNRRCVNTFPGSYCKCHIGFELQYISGRYCIDIDINECTMDSHTCSHANCFN 238
      |||
      |||
Db 1285 DLNPIVC-LSGCNCEHTKGSFICHCMGYSGK--KGTGCTDINECEIGAHNCGRHAVCTN 1341
      |||
      |||

QY 239 TQGSFKCKCKQGYKNGSLRCSAIPENS 265
      |||
      |||
Db 1342 TAGSFKCSKSPGMWIDGICKTCDLDECS 1368
      |||
      |||

```

fibrillin 1 precursor - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000

C:Accession: A47221; 154355; S17064; 159574; S17062; S62111; A34198

C:Corson, G.M.; Chaberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.

Genomics 17, 476-484, 1993

A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure

A:Reference number: A47221; MUID:94010947

A:Accession: A47221

A:Molecule type: mRNA

A:Residues: 1-337, 'T', 339-1029 <COR>

A:CROSS-references: GB:X63556

R:Peñalta, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Panglman, T.; Bo

Hum. Mol. Genet. 2, 961-968, 1993

A:Title: Genomic organization of the sequence coding for fibrillin, the defective gene

A:Reference number: 154355; MUID:93372860

A:Accession: 154355

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 132-3002 <LPR>

A:CROSS-references: GB:J19323; NID:9306745; PIDN:AB02036.1; PID:9306746

R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.

Nature 352, 334-337, 1991

A:Title: Partial sequence of a candidate gene for the Marfan syndrome.

A:Reference number: S17064; MUID:91304568

A:Accession: S17064

A:Molecule type: mRNA

A:Residues: 1030-3002 <MAS>

A:CROSS-references: EMBL:X63556

R:Dietz, H.C.; Valle, D.; Franccomano, C.A.; Kendzior, R.J.

Science 259, 680-683, 1993

A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.

A:Reference number: 159574; MUID:9157831

A:Accession: 159574

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 2217-2288, 'T', 2290-2325 <RES>

A:CROSS-references: GB:S54426; NID:9264860; PIDN:AB25244.1; PID:9264861

R:Rhee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras,

Nature 352, 330-334, 1991

A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different

A:Reference number: S17062; MUID:91304567

A:Accession: S17062

A:Molecule type: mRNA

A:Residues: 'VLVTYVFFLSYKMKL', 944-1444 <LEEI>

A:CROSS-references: EMBL:X65008; NID:931398; PIDN:CAB56534.1; PID:95924015

A:Accession: S62111

A:Molecule type: protein

A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>

R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.

J. Biol. Chem. 264, 21381-21385, 1989

A:Title: Connective tissue microfibrils. Isolation and characterization of three large

A:Reference number: A34198; MUID:90078246

A:Accession: A34198

A:Molecule type: protein

A:Residues: 565-575, 1890-1892, 'T', 1894-1900 <MAD>

C:Comment: Fibrillin is a major component of elastin-associated microfibrils.

C:Genetics:

A:Gene: GDB:FBN1

A:CROSS-references: GDB:127115; OMIM:134797; OMIM:154700

A:Map position: 15q21.1-15q21.1

A:Introns: 2236/1; 2258/1; 2297/1

A:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein

E:1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predict

E:1332-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MARC>

E:1332-1367/Domain: EGF homology <EGF>

E:1457-1492/Domain: EGF homology <EGF2>

E:2262-2295/Domain: EGF homology <EGF1>

RESULT 6	
A47221	
<hr/>	
Query Match	12.5%
Best Local Similarity	32.6%
Matches 87; Conservative	33; Mismatches 82; Indels 65; Gaps 11.
Score 383.5; DB 2;	Length 3002;
Pred. No. 1.2e-16;	

QY 39 GVCHYKTLACCGYMRNSKGVCEATCEPG-----CKFGEQV- 75
 1258 GVCH-----NTESSYRCECPGHHQLSPNISACIDINECELSAHLCPGRCV 1304
 Db 76 --GNKCKCFPGY---TGKTSQDVNECGMKPRPCQHRVCVTHGSKYKFCISLGHMLMDA 130
 1305 LIGRYOCACNPGYHSTPDLRCVLDIDECISINMGCEFTCTNSESSESCOPGFALMPD- 1363
 Db 131 TCVNSRTCAINCOYSCEDTE---EGPO-----CLCPSSGLRLAPNGRDCLDIDEC 178
 1364 ----QRSTIDID---ECEDNINICDGQCTNIPGEYRLC-YDGFMASEDMKTQVAVNEC 1415
 Db 179 ASGRVICYPNRCVNTFGSYCKCHIGELQYISGRYDCIDINECTMDSHTCSHHANCEN 238
 1416 DLNPNIC-LSGTCENTKSFICHCMDYSGK--KGKGTCTDINECEIGAHHCKGHAVC 1472
 QY 239 TQSGFKCKCKGKYGKNGLRCSAIPENS 265
 1473 TAGSFKCSGSPWIGDGIKCTDLDECS 1499

RESULT 7

A:Accession: A49457
 A:Species: Mus musculus (house mouse)
 C:Date: 03-May-1994 #sequence, revision 03-May-1994 #text, change 11-Jan-2000
 C:Accession: A49457; S74095
 R:Pan, T.C.; Sasaki, T.; Zhang, R.Z.; Faessler, R.; Timpl, R.; Chu, M.L.
 J. Cell Biol. 123, 1269-1277, 1993
 A:Title: Structure and expression of fibulin-2, a novel extracellular matrix protein with
 A:Reference number: A49457; MUID:94064787
 A:Accession: A49457
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1221 <PAN>
 A:Cross-references: GB:X75285; NID:9437046; PIDN:CA5340.1; PID:9437047
 R:Sasaki, T.; Mann, K.; Murphy, G.; Chu, M.L.; Timpl, R.
 Eur. J. Biochem. 240, 427-434, 1996
 A:Title: Different susceptibilities of fibulin-1 and fibulin-2 to cleavage by matrix met
 A:Reference number: S74094; MUID:96439073
 A:Accession: S74095
 A:Molecule type: protein
 A:Residues: 236-238, 'X', 240-247, 260-275, 336-344, 'L', 346-361, 405-426, 566-568, 'EM', 569-586
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: calcium binding; duplication; extracellular matrix; glycoprotein; homotrimer
 F:942-978/Domain: EGF homology <EG>

Query Match 12.5%; Score 382; DB 2; Length 1221;
 Best Local Similarity 28.5%; Pred. No. 6.3e-17;
 Matches 114; Conservative 39; Mismatches 139; Indels 108; Gaps 20;

QY 47 LACCGYMRNSKGV---VCEATCEPG---CKFGE-CV---GNKCKCFPGYT---GKTCS 92
 878 LVCGRGYHANESEGCYVNECEGVHRCGSGQLCYNLPGSYRCKCKGFPGRDAGRCRI 937
 Db 93 QDVNECGMKR-PRPCQHRVCVTHGSKYKFCISLGHMLMPAT-CVNSRTCAINCOYSCEDT 150
 938 -DVNECGVSPRLCQHTENPFGSYRCSAAGFLAAGKCEDEVECTRRCSQECANI 996
 QY 151 EEGPQCLPSSGLRLAPNGRDCLDIDECASGK-VICPYNRRCVNTFGSYCKCHIGFELQ 209
 997 YGSIQCYC-RQGYQLAEGRHTCTDIDECAGAGILCTE--RCVAVPGSYQACPEQGYTM 1053
 Db 210 YISGRYDCIDINECTMDSHTCSHHANCENFGSGFCK--CKCKGY-KGNGLRCSAIPENS 265
 1054 MANOR-SEKDLDECALGTHNCSAETCHNIGSFRLCFDPCPNVVRVSQTKCENTTQOD 1112
 QY 266 VKEVLRADGTIKDKIKLAKHNSKKKAKIKNVTPPTPTPTPKVNLQPFNYEIVSRG 325
 1113 ITECQTSFARI---THYQLNFQGTGLVPAHIFRIGPAP----- 1147
 Db 326 GNSHGKKGKNGEKKKEGLEDEKREKALKNDIEKSLAGDYF---FPVYNAGEF----- 377

Db 1148 -----AFAGDTISLTITGNGEEGFVTRRL 1172
 QY 378 ----GLIVORKALTSKLEHKLNLNISVDCSFNHHGICDCKQ 413
 Db 1173 NAYTVGSLSQR---SVLEPRDFALDVKML-----WRQ 1202

RESULT 8

MEGF6 protein - rat
 T13954
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence, revision 20-Sep-1999 #text, change 21-Jul-2000
 C:Accession: T13954
 R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motif
 A:Reference number: Z14126; MUID:98360089
 A:Accession: T13954
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1574 <NAK>
 A:Cross-references: EMBL:AB011532; NID:93449293; PIDN:BA432462.1; PID:93449294
 A:Experimental source: strain Sprague-Dawley; brain
 C:Genetics:
 A:Gene: MEGF6

Query Match 11.9%; Score 365.5; DB 2; Length 1574;
 Best Local Similarity 36.8%; Pred. No. 9.4e-16;
 Matches 93; Conservative 25; Mismatches 96; Indels 39; Gaps 13;

QY 34 SARQGVCHYGTAKLACCGYMR-----NSKGYCEATCEPGCKFGEQVGN 78
 92 SYROYATARTVFRCCGWSQKPGQGLSDVDCAASNGCEBPC--C--NTVGGF 145
 Db 79 KCRCPFGY---TGKTSQDVNECGMKPRPCQHRVCVTHGSKYKFCISLGHMLMPDA-TCV 133
 146 YCRCPFGYQLQDGDRTK-QDVDECAHNGGQHRVCVNTFGSYLCECKGFRLLHDGRTCL 204
 Db 134 NSRTCAIN--COYSC-EDTEEGPQCLPSSGLRLAPNGRDCLDIDECASGKVCIPYNR 190
 205 AISCCTLNGGCGQHCQVQLVYQHRCQCRPO-YQLQEDGRRCVRRSPCAEGNGGCMH--I 261
 QY 191 CVNTPFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCENFGSKCKCKQ 250
 262 QQLNLGLAHGCHCPGYL--AADRTCEVDCAALGAGCAH--GCLNTPGSKFCVCHAG 317
 Db 251 YK--GNGLRCSAI 261
 QY 318 YELGADGRQCYRI 330

RESULT 9

A55184
 fibulin-2 precursor - human
 N:Alternate names: protein DKFZ586A1519.1
 C:Species: Homo sapiens (man)
 C:Date: 27-Jan-1995 #sequence, revision 27-Jan-1995 #text, change 21-Jul-2000
 C:Accession: A55184; T08744
 R:Zhang, R.Z.; Pan, T.C.; Zhang, Z.Y.; Mattei, M.G.; Timpl, R.; Chu, M.L.
 Genomics 22, 425-430, 1994
 A:Title: Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping of the
 A:Reference number: A55184; MUID:95104855
 A:Accession: A55184
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1184 <ZHA>
 A:Cross-references: GB:X82494; NID:9575232; PIDN:CA57876.1; PID:9575233
 R:Mandut, R.; Heudner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z16471
 A:Accession: T08744
 A:Molecule type: mRNA

A:Molecule type: DNA
 A:Residues: 2504-2576,'E',2578-2611 <WHA2>
 C:Genetics:
 A:Gene: notch; opa
 A:Cross-references: FlyBase:FBgn0004647
 A:Map position: 8.96-9.36
 A:Interons: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
 C:Superfamily: notch protein; tandem repeat; transmembrane protein
 C:Keywords: differentiation; tandem repeat; transmembrane protein
 F:21743/Domain: transmembrane #status predicted <TM1>
 F:237-328/Domain: EGF homology <EGX1>
 F:530-561/Domain: EGF homology <EGF1>
 F:568-599/Domain: EGF homology <EGF>
 F:988-1019/Domain: EGF homology <EGX2>
 F:1064-1095/Domain: EGF homology <EGF3>
 F:1187-1218/Domain: EGF homology <EGX3>
 F:1746-1767/Domain: transmembrane #status predicted <TM2>
 F:1950-1982/Domain: ankyrin repeat homology <AN1>
 F:1983-2015/Domain: ankyrin repeat homology <AN2>
 F:1988-2004/Domain: transmembrane #status predicted <TM3>
 F:2017-2049/Domain: ankyrin repeat homology <AN3>
 F:2050-2082/Domain: ankyrin repeat homology <AN4>
 F:2083-2115/Domain: ankyrin repeat homology <AN5>
 F:2538-2568/Region: glutamine-rich
 F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 11.0%; Score 335.5; DB 1; Length 2703;
 Best Local Similarity 26.4%; Pred. No. 1.4e-13;
 Matches 96; Conservative 34; Mismatches 106; Indels 127; Gaps 17;

```

QY 39 GVCHYGT-----KLACCYGM-----RRSKGVCEAT-CEPCCKFGECVGPNK--CRCPGYTG 88
D 152 GTCLKLTLEETACANCIYTERGETKRLKLCASSPCRNATCTALAGSSFTSCSPRGFTG 211
QY 89 KTGSDVNECGMKPRPCOH--RCVNTGHSYKCFCLSG-----HMLMPATCVNSRT 137
D 212 DTCSYDIEEC--QSNPCYGGTGVNTHGSYOCMPTGYTKDKCDTKYKPCSPCQNGGI 269
QY 138 CAMINCOYSCEDTE--EGPOC-----LCPSSG-----LRLAPN--GRDCL 173
D 270 CRSGNLSTECKCPKGFESKNCENYDCLGHLCONGCTCIDISDYTCRCPNNTGRCQ 329
QY 174 -DIDECAS-----GKVICPYNRRCVN 193
D 330 DDVDECAQRDHPVCQNGATCTNTHGSYSCICVNGWAGLDCSNNTDDCAACFGATCID 389
QY 194 TFGSYTCKCHG-----FELAYISGRDCT-----D 219
D 390 GVGSPFYCQCTKGTGLCHLDDACTSNPCHADAICTSPINGSYACSCATGYKGVDCSED 449
QY 220 INECTMDSHTCSHANCENFTGSGFKCKQGYG-----NGLRCSAIPENSKEVLRAPG 274
D 450 IDECDQGS-PCENHICVNTPTGSTRNCSSGFTGPRCETNINECESHPCONEGSCLDDPG 508
QY 275 TIK 277
D 509 TFR 511

```

RESULT 15
 S34968
 fibulin, splice form D precursor - mouse
 N:Alternate names: basement-membrane protein BM-90; calcium-binding protein BM-90
 C:Species: Mus musculus (house mouse)
 C>Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 31-Jan-2000
 C:Accession: S34968; S36441; S13814
 R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
 Eur. J. Biochem. 215, 733-740, 1993
 A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-dependent
 A:Reference number: S34968; MUID:93358897
 A:Accession: S34968
 A:Molecule type: mRNA
 A:Residues: 1-705 <PAN>

R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timpl, R.; Chu, M.L.
 submitted to the EMBL Data Library, January 1993
 A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-
 A:Reference number: S36440
 A:Accession: S36441
 A:Molecule type: mRNA
 A:Residues: 1-39,'P',41-705 <PAN>
 A:Cross-references: EMBL:X10854; NID:g396820; PIDN:CA50207.1; PID:g396821
 A:Experimental source: cell-line F9 teratocarcinoma
 R:Kluge, M.; Mann, K.; Dziadek, M.; Timpl, R.
 Eur. J. Biochem. 193, 651-659, 1990
 A:Title: Characterization of a novel calcium-binding 90-kDa glycoprotein (BM-90) shar
 A:Reference number: S13814; MUID:91065369
 A:Accession: S13814

A:Molecule type: protein
 A:Residues: 28;31-49,'X',51-53,'XX',110-117;231-240,'X',242-243;339-362,'S',364-387;4
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular m
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-705/Product: fibulin, splice form D #status predicted <MAT>
 F:98;537;541/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.9%; Score 335; DB 2; Length 705;
 Best Local Similarity 34.1%; Pred. No. 3.9e-14;
 Matches 86; Conservative 31; Mismatches 87; Indels 48; Gaps 15;

```

QY 46 KLACCYGRRNSKGYC-----EATCEPCCKFGECVGNKCR-----CFPGY-----T 87
D 293 KLOCKSGFIQDALGNCIDINECLISAPCPYGGQTCINTGSGYTCQKNVPCNGRGYHNEE 352
QY 88 KGTCSODVNECGMKPRPC--QHRGVNTHGSYKCFCLSGHMLMPDATCVNSRTCAMIN--- 142
D 353 GTRC-VDVDECAPRAEPKCGKHHCINSPGSRCECKAFTY--DGI---SRTCDINECQ 406
QY 143 -----CQYSCDETEEGPOCLPSSGRLAPNGRDCIDIDECASGKYICPYNRRCVNFTG 196
D 407 RYPRRLCGHKCNENTPGSPHCSC-SAGFRLSYDGRSCEDVDNCLNS---PCSQECANYG 461
QY 197 SYTCKCHTGFELQYISGTYDCIDINECTMDS--HTCSHANCFTGSGFKKC-KQGYK- 252
D 462 SYQCYCRGYSGLSDVDG-VTCEIDIDECALPTGHTCSY--RCINIPGSGFQSCSPSSGYRL 518
QY 253 -GNGLRCSAIP 263
D 519 APNGRNCODIDE 530

```

Search completed: October 17, 2002, 16:01:56
 Job time : 32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 15:53:08 : Search time 14 Seconds
(without alignments)
1529.422 Million cell updates/sec

Title: US-09-687-860-24
Perfect score: 3060
Sequence: 1 MFLPMSLALPLLSWVAGF.....VDGVLLVSGLCPPSLSLSYDD 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	410.5	13.4	2907	1 FBN2_MOUSE	Q61555 mus musculus
2	409.5	13.4	2911	1 FBN2_HUMAN	P35556 homo sapien
3	387.5	12.7	2871	1 FBN1_BOVIN	P98133 bos taurus
4	385.5	12.6	2871	1 FBN1_MOUSE	Q61554 mus musculus
5	385.5	12.6	2871	1 FBN1_PIG	Q9TV36 sus scrofa
6	383.5	12.5	2871	1 FBN1_HUMAN	P35555 homo sapien
7	382	12.5	1221	1 FBL2_MOUSE	P37889 mus saplen
8	367	12.0	684	1 FBL1_CHICK	O73775 gallus gall
9	362.5	11.8	1184	1 FBL2_HUMAN	P98095 homo sapien
10	339	11.1	956	1 MTN2_MOUSE	O08746 mus musculus
11	338	11.0	703	1 FBL1_HUMAN	P23142 homo sapien
12	335.5	11.0	2703	1 NOTC_DROME	P07207 drosophila
13	335	10.9	703	1 FBL1_MOUSE	Q08879 mus musculus
14	332	10.8	2437	1 NOTC_BRARE	P46530 brachydanio
15	331	10.8	956	1 MTN2_HUMAN	O00339 homo sapien
16	330	10.8	712	1 FBL1_CAEEL	O77469 caenorhabd
17	329	10.8	1712	1 TGFBR_MOUSE	O00918 rattus norv
18	327.5	10.7	2318	1 MTN3_MOUSE	Q61982 mus musculus
19	325	10.6	2531	1 MTN1_RAT	Q07008 rattus norv
20	322.5	10.5	886	1 EMRI_HUMAN	Q14246 homo sapien
21	322.5	10.5	2444	1 MTN1_HUMAN	P46531 homo sapien
22	322	10.5	2531	1 MTN1_MOUSE	O01705 mus musculus
23	320	10.5	1394	1 TGFBR_HUMAN	P22064 homo sapien
24	315.5	10.3	1964	1 FBL1_MOUSE	Q9W419 mus musculus
25	315	10.3	1964	1 MTN4_MOUSE	P21655 mus musculus
26	313	10.2	443	1 FBL4_HUMAN	O95967 homo sapien
27	313	10.2	448	1 FBL5_HUMAN	O9ubk5 homo sapien
28	311.5	10.2	2524	1 NOTC_XENLA	P21783 xenopus lae
29	311	10.2	448	1 FBL5_RAT	O9W418 rattus norv
30	311	10.2	810	1 NEI1_RAT	O62919 rattus norv
31	311	10.2	835	1 CD97_HUMAN	P48960 homo sapien
32	310.5	10.1	810	1 NEI1_HUMAN	O92832 homo sapien
33	309.5	10.1	448	1 FBL5_MOUSE	Q9W419 mus musculus

34	308	10.1	443	1 FBL4_CRIGR	O55058 cricetus
35	294.5	9.6	2139	1 CRB_DROME	P10040 drosophila
36	292	9.5	714	1 DLL1_RAT	P97677 rattus norv
37	291	9.5	493	1 FBL3_HUMAN	Q12805 homo sapien
38	290.5	9.5	931	1 EMRI_MOUSE	O61568 mus musculus
39	290	9.5	493	1 FBL3_RAT	O35568 rattus norv
40	287.5	9.4	722	1 DLL1_MOUSE	O61483 mus musculus
41	279.5	9.1	816	1 NEI2_MOUSE	O62918 mus musculus
42	279	9.1	723	1 DLL1_RAT	O00548 rattus norv
43	279	9.1	816	1 NEI1_HUMAN	O90827 homo sapien
44	278	9.1	816	1 NEI2_CHICK	O61220 gallus gall
45	273	8.9	816	1 NEI2_HUMAN	O99435 homo sapien

ALIGNMENTS

```

RESULT 1
FBN2_MOUSE STANDARD: PRT: 2907 AA.
AC 061555; 063957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Fibrillin 2 precursor.
GN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9526370; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.;
RT "Developmental expression of fibrillin genes suggests heterogeneity
RT of extracellular microfibrils."
RL J. Cell Biol. 129:1165-1176(1995).
RN [2]
RP SEQUENCE OF 210-317 FROM N.A.
RX MEDLINE=94140368; PubMed=8307578;
RA Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
RA Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on
RT mouse chromosomes 2 and 18."
RL Genomics 18:667-672(1993).
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L39790; AA74908.1; -
DR EMBL: S69359; AAC60685.1; -
DR HSSP: P35555; IEMN.
DR MGD: MGI:95490; Fbn2.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR PRINTS: PR00010; EGFBLD.

```

[illegible]

FT	DISU.FID	280	292	BY	SIMILARITY
FT	DISU.FID	287	301	BY	SIMILARITY
FT	DISU.FID	303	316	BY	SIMILARITY
FT	DISU.FID	322	334	BY	SIMILARITY
FT	DISU.FID	329	343	BY	SIMILARITY
FT	DISU.FID	345	358	BY	SIMILARITY
FT	DISU.FID	491	503	BY	SIMILARITY
FT	DISU.FID	498	512	BY	SIMILARITY
FT	DISU.FID	514	526	BY	SIMILARITY
FT	DISU.FID	532	542	BY	SIMILARITY
FT	DISU.FID	537	551	BY	SIMILARITY
FT	DISU.FID	553	566	BY	SIMILARITY
FT	DISU.FID	572	584	BY	SIMILARITY
FT	DISU.FID	579	593	BY	SIMILARITY
FT	DISU.FID	595	608	BY	SIMILARITY
FT	DISU.FID	614	625	BY	SIMILARITY
FT	DISU.FID	620	634	BY	SIMILARITY
FT	DISU.FID	636	649	BY	SIMILARITY
FT	DISU.FID	655	666	BY	SIMILARITY
FT	DISU.FID	661	675	BY	SIMILARITY
FT	DISU.FID	677	690	BY	SIMILARITY
FT	DISU.FID	765	777	BY	SIMILARITY
FT	DISU.FID	772	786	BY	SIMILARITY
FT	DISU.FID	788	801	BY	SIMILARITY
FT	DISU.FID	807	819	BY	SIMILARITY
FT	DISU.FID	814	828	BY	SIMILARITY
FT	DISU.FID	830	843	BY	SIMILARITY
FT	DISU.FID	849	859	BY	SIMILARITY
FT	DISU.FID	854	868	BY	SIMILARITY
FT	DISU.FID	870	883	BY	SIMILARITY
FT	DISU.FID	952	964	BY	SIMILARITY
FT	DISU.FID	959	973	BY	SIMILARITY
FT	DISU.FID	975	988	BY	SIMILARITY
FT	DISU.FID	1070	1082	BY	SIMILARITY
FT	DISU.FID	1077	1091	BY	SIMILARITY
FT	DISU.FID	1093	1106	BY	SIMILARITY
FT	DISU.FID	1112	1124	BY	SIMILARITY
FT	DISU.FID	1119	1133	BY	SIMILARITY
FT	DISU.FID	1135	1149	BY	SIMILARITY
FT	DISU.FID	1155	1167	BY	SIMILARITY
FT	DISU.FID	1162	1176	BY	SIMILARITY
FT	DISU.FID	1178	1191	BY	SIMILARITY
FT	DISU.FID	1197	1209	BY	SIMILARITY
FT	DISU.FID	1204	1218	BY	SIMILARITY
FT	DISU.FID	1220	1233	BY	SIMILARITY
FT	DISU.FID	1239	1250	BY	SIMILARITY
FT	DISU.FID	1246	1259	BY	SIMILARITY
FT	DISU.FID	1261	1274	BY	SIMILARITY
FT	DISU.FID	1280	1292	BY	SIMILARITY
FT	DISU.FID	1287	1301	BY	SIMILARITY
FT	DISU.FID	1303	1316	BY	SIMILARITY
FT	DISU.FID	1322	1334	BY	SIMILARITY
FT	DISU.FID	1329	1343	BY	SIMILARITY
FT	DISU.FID	1345	1358	BY	SIMILARITY
FT	DISU.FID	1364	1377	BY	SIMILARITY
FT	DISU.FID	1371	1386	BY	SIMILARITY
FT	DISU.FID	1388	1399	BY	SIMILARITY
FT	DISU.FID	1405	1418	BY	SIMILARITY
FT	DISU.FID	1412	1427	BY	SIMILARITY
FT	DISU.FID	1429	1440	BY	SIMILARITY
FT	DISU.FID	1446	1458	BY	SIMILARITY
FT	DISU.FID	1453	1467	BY	SIMILARITY
FT	DISU.FID	1468	1482	BY	SIMILARITY
FT	DISU.FID	1488	1499	BY	SIMILARITY
FT	DISU.FID	1494	1508	BY	SIMILARITY
FT	DISU.FID	1510	1523	BY	SIMILARITY
FT	DISU.FID	1529	1540	BY	SIMILARITY
FT	DISU.FID	1535	1549	BY	SIMILARITY
FT	DISU.FID	1551	1564	BY	SIMILARITY
FT	DISU.FID	1647	1659	BY	SIMILARITY
FT	DISU.FID	1654	1668	BY	SIMILARITY
FT	DISU.FID	1670	1683	BY	SIMILARITY
FT	DISU.FID	1689	1701	BY	SIMILARITY

```

RL Hum. Genet. 103:22-28(1998).
RN [5]
RX VARIANTS CCA PHE-1141 AND TRP-1252.
RA MEDLINE=20259236; PubMed=10797416;
RA Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RT Godfrey M.;
RT "Two novel fibrillin-2 mutations in congenital contractural
arachnoidecty";
RL Am. J. Med. Genet. 92:7-12(2000)
CC -1 FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE
CC THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.
CC -1 DISEASE DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL
CC ARACHNOIDECTY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS
CC PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT EFFECT THE
CC AORTA AND THE EYES.
CC -1 SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1 SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U03272; AAA18950.1; -
DR EMBL: X62009; - NOT_ANNOTATED_CDS.
DR PIR: S17063; SI7063.
DR PIR: S31101; S31101.
DR HSSP: P35555; IEMN.
DR MM: 121050; -
DR InterPro: IPRO00152; Asx_hydroxy1.
DR InterPro: IPRO00561; EGF-like.
DR InterPro: IPRO01881; EGF_Ca.
DR InterPro: IPRO01438; EGF-II.
DR InterPro: IPRO02212; TB.
DR Pfam: PF00008; EGF; 45.
DR Pfam: PF00683; TB; 9.
DR PRINTS: PR00010; EGFBLOOD.
DR SMART: SM00179; EGF_CA; 43.
DR SMART: SM00001; EGF_Like; 3.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 37.
DR PROSITE: PS01187; EGF_CA; 43.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family; Disease mutation; Polymorphism.
FT SIGNAL 1
FT CHAIN 29
FT DOMAIN 111
FT DOMAIN 142
FT DOMAIN 145
FT DOMAIN 176
FT DOMAIN 207
FT DOMAIN 275
FT DOMAIN 317
FT DOMAIN 358
FT REPEAT 359
FT DOMAIN 425
FT REPEAT 493
FT DOMAIN 533
FT DOMAIN 534
FT DOMAIN 574
FT DOMAIN 616
FT DOMAIN 657
FT DOMAIN 697
FT REPEAT 767
FT DOMAIN 808
FT DOMAIN 809
FT DOMAIN 851
FT DOMAIN 890
FT REPEAT 954
FT REPEAT 996
FT DOMAIN 1071
FT DOMAIN 1072
FT DOMAIN 1113
FT DOMAIN 1114
FT DOMAIN 1157
FT DOMAIN 1199
FT DOMAIN 1240
FT FIBRILLIN 2.
FT EGF-LIKE 1, NON-CALCIUM BINDING.
FT EGF-LIKE 2, NON-CALCIUM BINDING.
FT EGF-LIKE 3, NON-CALCIUM BINDING.
FT EGF-LIKE 4, CALCIUM-BINDING.
FT EGF-LIKE 5, CALCIUM-BINDING.
FT TGFBP 1.
FT EGF-LIKE 6, NON-CALCIUM BINDING.
FT EGF-LIKE 7, CALCIUM-BINDING.
FT EGF-LIKE 8, CALCIUM-BINDING.
FT EGF-LIKE 9, CALCIUM-BINDING.
FT EGF-LIKE 10, CALCIUM-BINDING.
FT TGFBP 2.
FT EGF-LIKE 11, CALCIUM-BINDING.
FT EGF-LIKE 12, CALCIUM-BINDING.
FT EGF-LIKE 13, CALCIUM-BINDING.
FT EGF-LIKE 14, CALCIUM-BINDING.
FT TGFBP 3.
FT EGF-LIKE 15, CALCIUM-BINDING.
FT EGF-LIKE 16, CALCIUM-BINDING.
FT EGF-LIKE 17, CALCIUM-BINDING.
FT EGF-LIKE 18, CALCIUM-BINDING.
```

FT	DOMAIN	1281	1321	EGF-LIKE 19,	CALCIUM-BINDING
FT	DOMAIN	1282	1323	EGF-LIKE 20,	CALCIUM-BINDING
FT	DOMAIN	1324	1365	EGF-LIKE 21,	CALCIUM-BINDING
FT	DOMAIN	1366	1406	EGF-LIKE 22,	CALCIUM-BINDING
FT	DOMAIN	1407	1447	EGF-LIKE 23,	CALCIUM-BINDING
FT	DOMAIN	1448	1489	EGF-LIKE 24,	CALCIUM-BINDING
FT	DOMAIN	1490	1530	EGF-LIKE 25,	CALCIUM-BINDING
FT	DOMAIN	1531	1571	EGF-LIKE 26,	CALCIUM-BINDING
FT	REPEAT	1572	1648	EGF-LIKE 27,	CALCIUM-BINDING
FT	DOMAIN	1649	1690	EGF-LIKE 28,	CALCIUM-BINDING
FT	DOMAIN	1691	1732	EGF-LIKE 29,	CALCIUM-BINDING
FT	REPEAT	1733	1806	EGF-LIKE 30,	CALCIUM-BINDING
FT	DOMAIN	1807	1848	EGF-LIKE 31,	CALCIUM-BINDING
FT	DOMAIN	1849	1890	EGF-LIKE 32,	CALCIUM-BINDING
FT	DOMAIN	1891	1932	EGF-LIKE 33,	CALCIUM-BINDING
FT	DOMAIN	1933	1971	EGF-LIKE 34,	CALCIUM-BINDING
FT	DOMAIN	1972	2014	EGF-LIKE 35,	CALCIUM-BINDING
FT	DOMAIN	2015	2054	EGF-LIKE 36,	CALCIUM-BINDING
FT	DOMAIN	2055	2096	EGF-LIKE 37,	CALCIUM-BINDING
FT	REPEAT	2097	2169	EGF-LIKE 38,	CALCIUM-BINDING
FT	DOMAIN	2170	2211	EGF-LIKE 39,	CALCIUM-BINDING
FT	DOMAIN	2212	2251	EGF-LIKE 40,	CALCIUM-BINDING
FT	DOMAIN	2252	2292	EGF-LIKE 41,	CALCIUM-BINDING
FT	DOMAIN	2293	2336	EGF-LIKE 42,	CALCIUM-BINDING
FT	DOMAIN	2337	2378	EGF-LIKE 43,	CALCIUM-BINDING
FT	REPEAT	2379	2447	EGF-LIKE 44,	CALCIUM-BINDING
FT	DOMAIN	2448	2489	EGF-LIKE 45,	CALCIUM-BINDING
FT	DOMAIN	2490	2530	EGF-LIKE 46,	CALCIUM-BINDING
FT	DOMAIN	2531	2569	EGF-LIKE 47,	CALCIUM-BINDING
FT	DOMAIN	2570	2612	EGF-LIKE 48,	CALCIUM-BINDING
FT	DOMAIN	2613	2652	EGF-LIKE 49,	CALCIUM-BINDING
FT	DOMAIN	2653	2693	EGF-LIKE 50,	CALCIUM-BINDING
FT	DOMAIN	2694	2733	EGF-LIKE 51,	CALCIUM-BINDING
FT	DISULFID	115	124	BY SIMILARITY.	
FT	DISULFID	119	130	BY SIMILARITY.	
FT	DISULFID	132	141	BY SIMILARITY.	
FT	DISULFID	149	159	BY SIMILARITY.	
FT	DISULFID	153	164	BY SIMILARITY.	
FT	DISULFID	166	175	BY SIMILARITY.	
FT	DISULFID	180	190	BY SIMILARITY.	
FT	DISULFID	184	195	BY SIMILARITY.	
FT	DISULFID	197	206	BY SIMILARITY.	
FT	DISULFID	279	291	BY SIMILARITY.	
FT	DISULFID	286	300	BY SIMILARITY.	
FT	DISULFID	302	315	BY SIMILARITY.	
FT	DISULFID	321	333	BY SIMILARITY.	
FT	DISULFID	328	342	BY SIMILARITY.	
FT	DISULFID	344	359	BY SIMILARITY.	
FT	DISULFID	497	509	BY SIMILARITY.	
FT	DISULFID	504	518	BY SIMILARITY.	
FT	DISULFID	520	532	BY SIMILARITY.	
FT	DISULFID	538	548	BY SIMILARITY.	
FT	DISULFID	543	557	BY SIMILARITY.	
FT	DISULFID	559	572	BY SIMILARITY.	
FT	DISULFID	578	590	BY SIMILARITY.	
FT	DISULFID	585	599	BY SIMILARITY.	
FT	DISULFID	601	614	BY SIMILARITY.	
FT	DISULFID	620	631	BY SIMILARITY.	
FT	DISULFID	626	640	BY SIMILARITY.	
FT	DISULFID	642	655	BY SIMILARITY.	
FT	DISULFID	661	672	BY SIMILARITY.	
FT	DISULFID	667	681	BY SIMILARITY.	
FT	DISULFID	683	696	BY SIMILARITY.	
FT	DISULFID	771	783	BY SIMILARITY.	
FT	DISULFID	778	792	BY SIMILARITY.	
FT	DISULFID	794	807	BY SIMILARITY.	
FT	DISULFID	813	825	BY SIMILARITY.	
FT	DISULFID	820	834	BY SIMILARITY.	
FT	DISULFID	836	849	BY SIMILARITY.	
FT	DISULFID	855	865	BY SIMILARITY.	
FT	DISULFID	860	874	BY SIMILARITY.	
FT	DISULFID	876	888	BY SIMILARITY.	
FT	DISULFID	958	970	BY SIMILARITY.	

FT	DISULFID	965	979	BY SIMILARITY.
FT	DISULFID	981	994	BY SIMILARITY.
FT	DISULFID	1076	1088	BY SIMILARITY.
FT	DISULFID	1083	1097	BY SIMILARITY.
FT	DISULFID	1099	1112	BY SIMILARITY.
FT	DISULFID	1118	1130	BY SIMILARITY.
FT	DISULFID	1125	1139	BY SIMILARITY.
FT	DISULFID	1141	1155	BY SIMILARITY.
FT	DISULFID	1161	1173	BY SIMILARITY.
FT	DISULFID	1168	1182	BY SIMILARITY.
FT	DISULFID	1184	1197	BY SIMILARITY.
FT	DISULFID	1203	1215	BY SIMILARITY.
FT	DISULFID	1210	1224	BY SIMILARITY.
Query Match 13.4%; Score 409.5; DB 1; Length 2911;				
Best Local Similarity 28.6%; Pred. No. 7.6e-21;				
Matches 96; Conservative 37; Mismatches 92; Indels 111; Gaps 13;				
QY	38	PGCVHGYGT-----KLACCYGW-----RNSKGCVCATCEBPQ-CKRGECV--GPNKC	80	
DB	1080	PMCTGYGCKRNTIGSFKCKCNSGFALDMERKCTDIDECRISPDLCGSGICVNTFGSPECC	1139	
QY	81	RECFPY-----TGTCGQ-----D	94	
DB	1140	EEFEGEESGFMKKMCDIDGECERNPLCRGTCVNTBSSFCDCPLHEHSPEDECDV	1199	
QY	95	VNEGCKMRPCOH-RCVNTHGSGKCFCLSGHMLMPAT-CVNSRTCAMIN--CQYSCEDT	150	
DB	1200	INEGSLDNLGRNCNCVNMIGTQCSCNPGYATPDROGCTDIDECIMINGGCDTQCTNS	1259	
QY	151	EEGPGCLPSSSLRLAPNRGDCLDIDECASGAVTC-----PNNRCV-----	192	
DB	1260	EGSYRCSG-SEGVALMPGRSCADIDEEENNDIDGQCTINIPERYCLCYDGFMAAMD	1318	
QY	193	-----NFGSYCYCKHIGFELQYISGRYDCIDINECTMDSHT	229	
DB	1319	KMTCIDVNECDLNSINICMFGECENTKGSFICHCQIGYSVK--KGTGGCTDVECEIGAHN	1376	
QY	230	CSHANCFTNGSGFCKCKQGYKGNLRCSPAIPENS	265	
DB	1377	CDMHASCLNIPGSFSCREGWINGIRICIDIDECIS	1412	
RESULT 3				
FBN1_BOVIN				
ID	FBN1_BOVIN	STANDARD:	PRT:	2871 AA.
AC	P98133:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Fibrillin 1 precursor (MP340).			
GN	FBN1.			
OS	Bos taurus (bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OX	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin;			
RX	MEDLINE=95137597; PubMed=7835900;			
RT	Tilstra D.J., Potter K.A., Byers P.H.;			
RT	"Sequence of the coding region of the bovine fibrillin CDNA and			
RT	localization to bovine chromosome 10.";			
RL	Genomics 23:480-485(1994).			
RN	[2]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=96132851; PubMed=8557636;			
RA	Gibson M.A., Hatzinikola G., Kumaratille J.S., Sandberg L.B.,			
RA	Nicholl J.K., Sutherland G.R., Cleary E.G.;			
RT	"Further characterization of proteins associated with elastic fiber			
RT	microfibrils including the molecular cloning of MAGP-2 (MP25).";			
RL	J. Biol. Chem. 271:1096-1103(1996)			

CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PPM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS.
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; L28748; AAA74122.1; -.
DR HSSP; P35555; IABJ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR002212; TB.
DR Pfam; PF00008; EGF; 46.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PR00010; EGFBLD.
DR SMART; SM00179; EGF_CA; 42.
DR SMART; SM00001; EGF_like; 4.
DR PROSITE; PS00010; Asx_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 38.
DR PROSITE; PS01187; EGF_CA; 45.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
KW SIGNAL 1 27
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.
FT DOMAIN 246 287 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.
FT REPEAT 330 390 TGFBP 1.
FT DOMAIN 392 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6, NON-CALCIUM BINDING.
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 654 722 TGFBP 2.
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 952 1027 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1528 1605 TGFBP 4.
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1689 1765 TGFBP 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.

FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2055 2126 TGFBP 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT REPEAT 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2401 TGFBP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2607 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2647 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 792 790 BY SIMILARITY.
FT DISULFID 805 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.

FT DISULFID 1182 1195 BY SIMILARITY.
 FT DISULFID 1201 1212 BY SIMILARITY.
 FT DISULFID 1208 1221 BY SIMILARITY.
 FT DISULFID 1223 1236 BY SIMILARITY.
 FT DISULFID 1242 1254 BY SIMILARITY.
 FT DISULFID 1249 1263 BY SIMILARITY.
 FT DISULFID 1265 1278 BY SIMILARITY.
 FT DISULFID 1284 1296 BY SIMILARITY.
 FT DISULFID 1291 1305 BY SIMILARITY.
 FT DISULFID 1307 1320 BY SIMILARITY.
 FT DISULFID 1326 1339 BY SIMILARITY.
 FT DISULFID 1333 1348 BY SIMILARITY.
 FT DISULFID 1350 1361 BY SIMILARITY.
 FT DISULFID 1367 1380 BY SIMILARITY.
 FT DISULFID 1374 1389 BY SIMILARITY.
 FT DISULFID 1391 1402 BY SIMILARITY.
 FT DISULFID 1408 1420 BY SIMILARITY.
 FT DISULFID 1415 1429 BY SIMILARITY.
 FT DISULFID 1431 1444 BY SIMILARITY.
 FT DISULFID 1450 1461 BY SIMILARITY.
 FT DISULFID 1456 1470 BY SIMILARITY.
 FT DISULFID 1472 1485 BY SIMILARITY.
 FT DISULFID 1491 1502 BY SIMILARITY.
 FT DISULFID 1497 1511 BY SIMILARITY.
 FT DISULFID 1513 1526 BY SIMILARITY.
 FT DISULFID 1610 1622 BY SIMILARITY.

Query Match 12.7%; Score 387.5; DB 1; Length 2871;
 Best Local Similarity 30.6%; Pred. No. 2.6e-19;
 Matches 89; Conservative 34; Mismatches 91; Indels 77; Gaps 10;

QY 49 CCYGRNRKSKVCATCPG-----CFEGCV--GPNKRC 82
 DB 1081 CGRQCVPPTPDPECKCEGESGFMNKNCDIDECORPDLICGVCLMTGESYREC 1140
 QY 83 FPGY--TGKCSODVNECGMKPRPCOH-RCYNTHSGYKFCCLSGHMLMPAT-CVNSRT 137
 DB 1141 PPGHDLANINACDINDINCELSAHLCPHGRVNLIGKQCACNPGYHSTPRLFCVDD 1200
 QY 138 CAMIN--CQYSEPTDEEGPQCLPSSGLRLAPNGHDCIDIDECASGVKIC-----PY 187
 DB 1201 CSIMNGGCTFCTNSESSEYSC-OPGFALMPDQSTCIDIDECEDNPNICDGGCTNIPG 1259
 QY 188 NRRCV-----NTGSYCKKHIFELQYISGR 214
 DB 1260 EYRLCYDGFMASEDMKTCVDVNECDLNPNICLSGTCENTGSPFICHCMDGYSGR--RGK 1317
 QY 215 YDCIDINECTDMSHTCSHANCFTPGSFCKCKGQYKNGILRCSAIPENS 265
 DB 1318 TGCIDINCEIGAHNCDRAVCTNAGSFCKSCSPGWIIGDIKTDLECS 1368

RESULT 4
 FBNI_MOUSE STANDARD; PRT; 2871 AA.
 AC Q61554; 060826;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Fibrillin 1 precursor.
 GN FBNI OR FBNI-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 1;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95130561; PubMed=7829516;
 RA Yin W., Germiller J., Sanguinetti C., Smiley E., Pangillian T.,
 RA Pereira L., Ramirez F., Bonadio J.;
 RT "Primary structure and developmental expression of Fbn-1, the mouse
 fibrillin gene";
 RL J. Biol. Chem. 270:11798-1806(1995).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Kidney;
 RA Ota K., Kumar A., Mada J., Kanwar Y.S.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
 CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
 CC MICROFIBRILS (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
 CC EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: L29454; AAA56840.1; -;
 DR EMBL: U22493; AAA64217.1; -;
 DR HSSP: P35353; IAPJ.
 DR MGD: MGI:95489; Fbn1.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR002212; TB.
 DR Pfam: PF00008; EGF; 46.
 DR Pfam: PF00683; TB; 9.
 DR SMART: SM00179; EGF_CA; 42.
 DR SMART: SM00001; EGF_Like; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 43.
 DR PROSITE: PS00023; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 38.
 DR PROSITE: PS01187; EGF_CA; 45.
 DR KX Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
 DR Repeat; signal; Multigene family.
 FT SIGNAL 1 27
 FT CHAIN 28 2871
 FT DOMAIN 81 112 EGF-Like 1, NON-CALCIUM BINDING.
 FT DOMAIN 115 146 EGF-Like 2, NON-CALCIUM BINDING.
 FT DOMAIN 147 178 EGF-Like 3, NON-CALCIUM BINDING.
 FT DOMAIN 246 287 EGF-Like 4, CALCIUM-BINDING.
 FT DOMAIN 288 329 EGF-Like 5, CALCIUM-BINDING.
 FT DOMAIN 330 401 TGFBRP 1.
 FT DOMAIN 402 446 PRO-RICH.
 FT DOMAIN 449 489 EGF-Like 6, NON-CALCIUM BINDING.
 FT DOMAIN 490 529 EGF-Like 7, CALCIUM-BINDING.
 FT DOMAIN 530 571 EGF-Like 8, CALCIUM-BINDING.
 FT DOMAIN 572 612 EGF-Like 9, CALCIUM-BINDING.
 FT DOMAIN 613 653 EGF-Like 10, CALCIUM-BINDING.
 FT DOMAIN 656 721 TGFBRP 2.
 FT DOMAIN 723 764 EGF-Like 11, CALCIUM-BINDING.
 FT DOMAIN 765 806 EGF-Like 12, CALCIUM-BINDING.
 FT DOMAIN 807 846 EGF-Like 13, CALCIUM-BINDING.
 FT DOMAIN 910 951 EGF-Like 14, CALCIUM-BINDING.
 FT DOMAIN 952 1018 TGFBRP 3.
 FT DOMAIN 1028 1069 EGF-Like 15, CALCIUM-BINDING.
 FT DOMAIN 1070 1112 EGF-Like 16, CALCIUM-BINDING.
 FT DOMAIN 1113 1154 EGF-Like 17, CALCIUM-BINDING.
 FT DOMAIN 1155 1196 EGF-Like 18, CALCIUM-BINDING.
 FT DOMAIN 1197 1237 EGF-Like 19, CALCIUM-BINDING.
 FT DOMAIN 1238 1279 EGF-Like 20, CALCIUM-BINDING.
 FT DOMAIN 1280 1321 EGF-Like 21, CALCIUM-BINDING.
 FT DOMAIN 1321 1362 EGF-Like 22, CALCIUM-BINDING.
 FT DOMAIN 1362 1403 EGF-Like 23, CALCIUM-BINDING.
 FT DOMAIN 1403 1445 EGF-Like 24, CALCIUM-BINDING.
 FT DOMAIN 1445 1486 EGF-Like 25, CALCIUM-BINDING.
 FT DOMAIN 1487 1527 EGF-Like 26, CALCIUM-BINDING.

FT	DOMAIN	1528	1599	TGFBP 4.
FT	DOMAIN	1606	1647	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1648	1688	EGF-LIKE 28, CALCIUM-BINDING.
FT	DOMAIN	1689	1758	TGFBP 5.
FT	DOMAIN	1766	1807	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1808	1848	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1849	1890	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1891	1929	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1930	1972	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	1973	2012	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2013	2054	EGF-LIKE 35, CALCIUM-BINDING.
FT	DOMAIN	2055	2121	TGFBP 6.
FT	DOMAIN	2127	2165	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2166	2205	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2206	2246	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCIUM-BINDING.
FT	DOMAIN	2333	2400	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING.
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.

FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.
FT	DISULFID	1326	1339	BY SIMILARITY.
FT	DISULFID	1333	1348	BY SIMILARITY.
FT	DISULFID	1350	1361	BY SIMILARITY.
FT	DISULFID	1367	1380	BY SIMILARITY.
FT	DISULFID	1374	1389	BY SIMILARITY.
FT	DISULFID	1391	1402	BY SIMILARITY.
FT	DISULFID	1408	1420	BY SIMILARITY.
FT	DISULFID	1415	1429	BY SIMILARITY.
FT	DISULFID	1431	1444	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1456	1470	BY SIMILARITY.
FT	DISULFID	1472	1485	BY SIMILARITY.
FT	DISULFID	1491	1502	BY SIMILARITY.
FT	DISULFID	1497	1511	BY SIMILARITY.
FT	DISULFID	1513	1526	BY SIMILARITY.
FT	DISULFID	1610	1622	BY SIMILARITY.
FT	DISULFID	1617	1631	BY SIMILARITY.
FT	DISULFID	1633	1646	BY SIMILARITY.
FT	DISULFID	1652	1663	BY SIMILARITY.
FT	DISULFID	1658	1672	BY SIMILARITY.

Query Match 12.6% Score 385.5; DB 1; Length 2871;
 Best Local Similarity 32.2%, Pred. No. 3,7e-19;
 Matches 86; Conservative 34; Mismatches 82; Indels 65; Gaps 11;

QY	39	GVCHYGRKLACCYGMRNSKGVCEATCEPG-----CKRGECV-75
DB	1127	GICH-----WTEGYRCECPGHQLSPNISACIDINECELSANICPHGRCVN-1173
QY	76	--GPNKRCQCPG---TKTCSODVNECGMKRPPQOHNCVMTFHSYKCFCLSGHMLMDA-130
DB	1174	LIGKYQACNPGYHPTHDRLFCVDDIECSIMNGCCTNSDGSYECSCOPGFALMPD-1232
QY	131	TCVNSRTCAMINCOYSCDTE---EGPQ-----CLCPSSGURLAPNGRDCIDIDEC-178
DB	1233	---QRSCSTD---QCEDNNICDGGCTNIPGEYRLC-YDGFMASEDMKTCVDVNEC-1284
QY	179	ASGKYICPNRCVNFESYCKCHIGFELYISGRYCIDINECTMDSHSCSHANCFN-238
DB	1285	DLNPNIC-LSGTCENTKCSFICHCDMGYSGR--KKGTCFTDINECEIGHANCGRAVCTN-1341
QY	239	TGSPFKCKQGYKGNGLRCSAIPENS-265
DB	1342	TAGSPFKCSPGWIGDGIKCTDLECS-1368
RESULT 5		
FBNI_PIG	STANDARD;	PRT: 2871 AA.
AC	Q9TV36;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Fibrillin 1 precursor.	
GN	FBNI.	
OS	Sus scrofa (pig).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	


```

FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
FT DISULFID 1674 1687 BY SIMILARITY.
FT DISULFID 1770 1782 BY SIMILARITY.
FT DISULFID 1777 1791 BY SIMILARITY.
FT DISULFID 1793 1806 BY SIMILARITY.
FT DISULFID 1812 1824 BY SIMILARITY.
FT DISULFID 1818 1833 BY SIMILARITY.
FT DISULFID 1835 1847 BY SIMILARITY.

```

Query Match 12.6%; Score 385.5; DB 1; Length 2871;
 Best Local Similarity 27.2%; Pred. No. 3.7e-19;
 Matches 91; Conservative 38; Mismatches 84; Indels 121; Gaps 11;

```

QY 49 CCYGMRRNSKVCBATCEPG-----CKFGECV--GPNKRCF 83
DB 1039 CTGKCKNTIGSFRCRDSGFALDSEERNCIDIDECHRISPDLCGRGCVNTPGDFECKD 1098
QY 84 PGTG-----KTSQDVNEGCMKRPQOHR-CVNFHSGYCFCLSGHMLMDAT----- 131
DB 1099 EGESGFMANKNC-MDIDECORPLCLRGCVLNTGSEYRCEPSGQMSFNISACIDIN 1157
QY 132 -----CVNSRGCAMN--COYCEDTEE 152
DB 1158 ECELSAHLCPHRCVNLIGYQRRARNPGYSTPRLFCVDDIDECISINMGCEFTCTNSEG 1217
QY 153 GPQCLPSSGRLAPNGRDCLIDECASGVYC-----PYNRRCV----- 192
DB 1218 SYEESC-QPFPALMPDRGCTDIDECEDNPNICOGGCTNIPGEYRCLCYDGFVASEDMK 1276
QY 193 -----NFGSYCKCHIGFEIQLYISGRYDCIDINECTMDSHFCS 231
DB 1277 TCVDVNECDLNPNICLSGTCFTGSGFICHCMDGYSK--KGKTCYDINECEIGAHCND 1334
QY 232 HHANCFNTOGSPFKCKCKGKYKGNGLRCSAIPENS 265
DB 1335 RHAVCTYTFAGSFNCSCSPWITGIDGICKTDLDECS 1368

```

```

RESULT 6
FBNI_HUMAN STANDARD; PRT; 2871 AA.
ID FBNI_HUMAN
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor.
GN FBNI OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangillan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
RN 12;
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta, and fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end.";
RL Genomics 17:476-484(1993).
RN 13;
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304567; PubMed=1852207;
RA Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
RN 14;
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;
RA Lee B., Godfrey M., Vitale E., Hori H., Mattel M.-G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister D.W.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes.";
RL Nature 352:330-334(1991).
RN 15;
RP CHARACTERIZATION.
RX MEDLINE=91317849; PubMed=1860873;
RA Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
RT "Purification and partial characterization of fibrillin, a cysteine-
RT rich structural component of connective tissue microfibrils.";
RL J. Biol. Chem. 266:14763-14770(1991).
RN 16;
RP STRUCTURE BY NMR OF 2054-2125.
RX MEDLINE=98031893; PubMed=9362480;
RA Yuan X., Downing A.K., Knott V., Handford P.A.;
RT "Solution structure of the transforming growth factor beta-binding
RT protein-like module, a domain associated with matrix fibrils.";
RL EMBO J. 16:6659-6666(1997).
RN 17;
RP STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE=96144829; PubMed=8568869;
RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;
RT "Calcium binding properties of an epidermal growth factor-like domain
RT pair from human fibrillin-1.";
RL J. Mol. Biol. 255:22-27(1996).
RN 18;
RP STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE=96222301; PubMed=8653794;
RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
RA Handford P.A.;
RT "Solution structure of a pair of calcium-binding epidermal growth

```



```

Oy      179 ASGKVCYCNRCVNTFFCSYCCKEGLXYISRYCVIDINECTMTSHNCANEN 238
         |::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1285 DLNPNIC-LSGTCENTTKTSFICHCDMGITSGR--KKGTCTDINECEIGHANCGRHACVTN 1341

Oy      239 TQGSFKCRCKOGYGNGDLRSCSAIPENS 265
         ||||||| |::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1342 TAGSFKCSCSPGWIGDGJIKTDTLDECS 1368


RESULT 7
FBIL2_MOUSE ID STANDARD; PRT; 1221 AA.
AC p37889; O9NWU12;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-Oct-2001 (Rel. 40, Last annotation update)
DE Fbibulin-2 precursor.
GN FBILN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-35.
RC TISSUE-Fibroblast;
RX MEDLINE=94064787; PubMed=8245130;
RA Pan T.-C.; Sasaki T.; Zhang R.-Z.; Paessler R.; Timpl R.; Chu M.-L.;
RT "Structure and expression of fibulin-2, a novel extracellular matrix protein with multiple EGF-like repeats and consensus motifs for calcium binding.";
RT J. Cell Biol. 123:1269-1277(1993);
[2]
RN RP SEQUENCE FROM N.A..
RX MEDLINE=99337686; PubMed=10406956;
RA Graessell S.; Sicot F.-X.; Gotta S.; Chu M.-L.;
RT "Mouse fibulin-2 gene. Complete exon-intron organization and promoter characterization.";
RL Eur. J. Biochem. 263:471-477(1999).
CC -I- FUNCTION: ITS BINDING TO FIBRONETIN AND SOME OTHER LIGANDS IS CALCIUM DEPENDENT.
CC -I- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -I- SUBCELLULAR LOCATION: Extracellular matrix.
CC -I- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF FIBULIN-2 DUE TO THE PRESENCE OR ABSENCE OF A SINGLE EGF-LIKE (3) REPEAT MAY ARISE BY ALTERNATIVE SPLICING.
CC -I- TISSUE SPECIFICITY: COMPONENT OF BOTH BASEMENT MEMBRANES AND OTHER CONNECTIVE TISSUES.
CC -I- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 11 EGF-LIKE DOMAINS.
-----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@sib-stb.ch).
-----
Cc EMBL; X75285; CAA53040.1; -.
DR EMBL; AF135253; AAD34456.1; JOINED.
DR EMBL; AF135259; AAD34456.1; JOINED.
DR EMBL; AF135240; AAD34456.1; JOINED.
DR EMBL; AF135241; AAD34456.1; JOINED.
DR EMBL; AF135242; AAD34456.1; JOINED.
DR EMBL; AF135243; AAD34456.1; JOINED.
DR EMBL; AF135244; AAD34456.1; JOINED.
DR EMBL; AF135245; AAD34456.1; JOINED.
DR EMBL; AF135246; AAD34456.1; JOINED.
DR EMBL; AF135247; AAD34456.1; JOINED.
DR EMBL; AF135248; AAD34456.1; JOINED.
DR EMBL; AF135249; AAD34456.1; JOINED.
DR EMBL; AF135250; AAD34456.1; JOINED
```

[illegible]

QT	SEQUENCE	684 AA:	75623 MM:	1638DVA094739199 CR664:
FT	CARBOHYD	510	524	BY SIMILARITY.
FT	CARBOHYD	530	543	BY SIMILARITY.
FT	DISULFID	537	552	BY SIMILARITY.
FT	DISULFID	557	569	BY SIMILARITY.
FT	CARBOHYD	96	96	N-LINKED (GLC/CNC . . .)
FT	CARBOHYD	536	536	N-LINKED (GLC/CNC . . .)
FT	CARBOHYD	540	540	N-LINKED (GLC/CNC . . .)
FT	CARBOHYD	540	540	N-LINKED (GLC/CNC . . .)
FT	CARBOHYD	638	638	N-LINKED (GLC/CNC . . .)

Query Match	12.0%;	Score 367;	DB 1;	Length 684;
Best Local Similarity	25.5%;	Pred. No. 1.5e-18;		
Matches 116;	Conservative 54;	Mismatches 129;	Indels 156;	Gaps 22

QY	38	PGVHYGKTLACCGYMRNSKSG-----VEATEPEPCFGE-CV---GPNKSC-----C	82
Db	284	PGSRCPKPLCKOMNGFQIDALGNCIDIDNECLSTMPP-CRPAQOICITINDGSIYTCORISPSC	342
QY	83	PPGJ-----TGKTCSDYNECGMKPRPC--OHRCVTHGSKYKCFCLSGHMLMPATCVNSR	136
Db	343	GRGHYNLEDGTRC-VDDDECSSSDQPCGEGHVACINGGNRCRCEKSGY-----SFVISR	396
QY	137	TCAMIN-----COYSCDPEEGPOCLCPSSGLPLAANGRCIDIDECASGKVIQPY	167
Db	397	TCIDINECRTPRGKCAHKCENTPGSIYTC-TWGEKLSDBGKCEDLINESS-----PC	451
QY	188	NRRCVNTPGSGYYCKCHIGFELYISG-----	213
Db	452	SOECANYYGSYQCYCRGFGOLDIDIGISCEIDICALPTGSHICFRFCINIPGSGQTCP	511
QY	214	-----RDCIDIDNECTMDSTGCSHANCFTQSPFC---KCKQGTKGNG-LRCSA	260
Db	512	STGYRLAPNARNCCDIDECVAETHNCSEFNTECFNIGGFRCLSLCEPCENYRKSGDTFCER	571
QY	261	IPESYKEVTLAPOTIKDIRIKLLAHKNSMKKAKIKNVPEPTPTPTPVNLQPFVYEE	320
Db	572	LPCHENNECGLP-----LRITYHLS-----PPTNIQVPTDIFRRGP	609
QY	321	IVSNGSGSHGCK-----GNEKKMKEGLEDEKREKALKNDIEERSLGDVFPFVYNE	373
Db	610	-----SNAVPDCKILLISIGNDCG-----FFTYKKV	636
QY	374	AGEFGLIVORKALTSKLEHNDIISYDCSP-NHG	407
Db	637	NHSGIYVMQNR-----TEPRDELITIQMLTRHG	667
RESULT 9			
FBL2_HUMAN	ID	FBL2_HUMAN	STANDARD; PRT; 1184 AA.
AC	P98095;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Fibulin-2 precursor.		
GN	FBLN2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	TISSUE=Fibroblast;		
RX	MEDLINE=95104855; PubMed=7806230;		
RA	Zhang R.-Z., Pan T.-C., Zhang Z.-Y., Mattei M.-G., Timpi R.,		
RA	Chu M.-L.;		
RT	*Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping		
RT	of the gene on human and mouse chromosomes.";		
RL	Genomics 22:425-430(1994).		
CC	-1- FUNCTION: ITS BINDING TO FIBRONCTIN AND SOME OTHER LIGANDS IS		
CC	CALCIUM DEPENDENT.		
CC	-1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.		
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX. COMPONENT OF BOTH		

[illegible]

FT DISULFID 821 836 BY SIMILARITY.
 FT DISULFID 843 856 BY SIMILARITY.
 FT DISULFID 862 875 BY SIMILARITY.
 FT DISULFID 869 884 BY SIMILARITY.
 FT DISULFID 886 899 BY SIMILARITY.
 FT DISULFID 905 917 BY SIMILARITY.
 FT DISULFID 913 926 BY SIMILARITY.
 FT DISULFID 928 941 BY SIMILARITY.
 FT DISULFID 947 956 BY SIMILARITY.
 FT DISULFID 952 965 BY SIMILARITY.
 FT DISULFID 967 980 BY SIMILARITY.
 FT DISULFID 986 998 BY SIMILARITY.
 FT DISULFID 994 1007 BY SIMILARITY.
 FT DISULFID 1009 1023 BY SIMILARITY.
 FT DISULFID 1029 1042 BY SIMILARITY.
 FT DISULFID 1036 1051 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT CARBOHYD 180 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1035 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1184 AA; 126543 MW; CAA8490A559EC5D CRC64;

Query Match 11.8%; Score 362.5; DB 1; Length 1184;
 Best Local Similarity 30.7%; Pred. No. 5,6e-18;
 Matches 107; Conservative 35; Mismatches 149; Indels 57; Gaps 17;

QY 21 GNAASARHGLLASAROPGVCHYGT-----KLACCYWRNRNSKG--VCEATCEPG 68
 DB 806 GNCVDINECTSLSEPCRFSCINTVSGYTCQRPMLICARGYHASDDACVDAVECEGT 865
 QY 69 ---GKPGCV---GPNKCRGCPGYT---GKTCSDVNECGMKP-RRCQRCVNTHSY 116
 DB 866 VHRGEGGVCHNLPGSYKCDKAGFQDAFGRGCI-DVNECWASBGRICQHTCENTLGSY 924
 QY 117 KCFLSGHMLPDA-TCVNSRTCAMINQYSCDETEEGPCLCPSSGRLAPNGRCDI 175
 DB 925 RGSASGFLAADKRCEDVNECEAQRSCQCANITYGYCYC-RQGQLADEGHTCDI 983
 QY 176 DECAAGK-VICPVNRRCVNFSGSYCKCHIGFELQYISGRYDCIDINECTMSHTCSHHA 234
 DB 984 DECAAGAGILCTF--RCILNPGSYOCACPEQGYMTANGR-SCKDVDECALGTNCSBAE 1040
 QY 235 NCFWQGSFKC---KCKQGY-KGNGLCRCSALPENSVEVLAPGTINDRIKILAHKNSM 290
 DB 1041 TCHNIGSFRLRCPCEPNYQVSKTCERTCHDELECONSAPRI--THYQLNFGTGL 1097
 QY 291 KKKAKIKNVPEPTPT-PTPKVNLQPFNYEIEIVSRGNSHGKKGKNEE 337
 DB 1098 LVPAHIFRIGPAPAFGTDTALNT-----IKGNEE 1127

RESULT 10
 MTN2_MOUSE
 ID MTN2_MOUSE STANDARD; PRT; 956 AA.
 AC 008746;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Matrilin-2 precursor.
 GN MATN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C; TISSUE=Limb;
 RX MEDLINE=97238863; PubMed=9083061;
 RA Deak F., Piecha D., Bachmati C., Paulsson M., Kiss I.;
 RT Primary structure and expression of matrilin-2, the closest relative
 of cartilage matrix protein within the von Willebrand factor type A-
 like module superfamily.";
 RT J. Biol. Chem. 272:9268-9274(1997).

CC -1- FUNCTION: INVOLVED IN MATRIX ASSEMBLY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: DETECTED IN A VARIETY OF ORGANS, INCLUDING
 CC CALVARIA, UTERUS, HEART, AND BRAIN, AS WELL AS FIBROBLAST AND
 CC OSTEOBLAST CELL LINES.
 CC -1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 WFPA DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U69262; AAC53163.1; --
 DR HSSP: P35555; IEMN.
 DR MGD: MGI:109613; Matn2.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-Like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR002035; WFPA.
 DR Pfam: PF00008; EGF; 10.
 DR Pfam: PF00092; Wfa; 2.
 DR PRINTS: PRO0453; WFPADOMAIN.
 DR SMART: SM00181; EGF; 10.
 DR SMART: SM00327; WFA; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 7.
 DR PROSITE: PS01186; EGF; 2; 9.
 DR PROSITE: PS50234; WFA; 2.
 DR EGF-Like domain; Signal; Glycoprotein; Repeat; Coiled coil.
 KW EGF-Like domain; Signal; Glycoprotein; Repeat; Coiled coil.
 FT CHAIN 1 23
 FT SIGNAL 1 23
 FT EFT 24 956
 FT EFT 57 232
 FT EFT 238 278
 FT EFT 279 319
 FT EFT 320 360
 FT EFT 361 401
 FT EFT 402 442
 FT EFT 443 483
 FT EFT 484 524
 FT EFT 525 565
 FT EFT 566 606
 FT EFT 607 647
 FT EFT 647 687
 FT EFT 687 727
 FT EFT 727 767
 FT EFT 767 807
 FT EFT 807 847
 FT EFT 847 887
 FT EFT 887 927
 FT EFT 927 967
 FT EFT 967 1007
 FT EFT 1007 1047
 FT EFT 1047 1087
 FT EFT 1087 1127
 FT EFT 1127 1167
 FT EFT 1167 1207
 FT EFT 1207 1247
 FT EFT 1247 1287
 FT EFT 1287 1327
 FT EFT 1327 1367
 FT EFT 1367 1407
 FT EFT 1407 1447
 FT EFT 1447 1487
 FT EFT 1487 1527
 FT EFT 1527 1567
 FT EFT 1567 1607
 FT EFT 1607 1647
 FT EFT 1647 1687
 FT EFT 1687 1727
 FT EFT 1727 1767
 FT EFT 1767 1807
 FT EFT 1807 1847
 FT EFT 1847 1887
 FT EFT 1887 1927
 FT EFT 1927 1967
 FT EFT 1967 2007
 FT EFT 2007 2047
 FT EFT 2047 2087
 FT EFT 2087 2127
 FT EFT 2127 2167
 FT EFT 2167 2207
 FT EFT 2207 2247
 FT EFT 2247 2287
 FT EFT 2287 2327
 FT EFT 2327 2367
 FT EFT 2367 2407
 FT EFT 2407 2447
 FT EFT 2447 2487
 FT EFT 2487 2527
 FT EFT 2527 2567
 FT EFT 2567 2607
 FT EFT 2607 2647
 FT EFT 2647 2687
 FT EFT 2687 2727
 FT EFT 2727 2767
 FT EFT 2767 2807
 FT EFT 2807 2847
 FT EFT 2847 2887
 FT EFT 2887 2927
 FT EFT 2927 2967
 FT EFT 2967 3007
 FT EFT 3007 3047
 FT EFT 3047 3087
 FT EFT 3087 3127
 FT EFT 3127 3167
 FT EFT 3167 3207
 FT EFT 3207 3247
 FT EFT 3247 3287
 FT EFT 3287 3327
 FT EFT 3327 3367
 FT EFT 3367 3407
 FT EFT 3407 3447
 FT EFT 3447 3487
 FT EFT 3487 3527
 FT EFT 3527 3567
 FT EFT 3567 3607
 FT EFT 3607 3647
 FT EFT 3647 3687
 FT EFT 3687 3727
 FT EFT 3727 3767
 FT EFT 3767 3807
 FT EFT 3807 3847
 FT EFT 3847 3887
 FT EFT 3887 3927
 FT EFT 3927 3967
 FT EFT 3967 4007
 FT EFT 4007 4047
 FT EFT 4047 4087
 FT EFT 4087 4127
 FT EFT 4127 4167
 FT EFT 4167 4207
 FT EFT 4207 4247
 FT EFT 4247 4287
 FT EFT 4287 4327
 FT EFT 4327 4367
 FT EFT 4367 4407
 FT EFT 4407 4447
 FT EFT 4447 4487
 FT EFT 4487 4527
 FT EFT 4527 4567
 FT EFT 4567 4607
 FT EFT 4607 4647
 FT EFT 4647 4687
 FT EFT 4687 4727
 FT EFT 4727 4767
 FT EFT 4767 4807
 FT EFT 4807 4847
 FT EFT 4847 4887
 FT EFT 4887 4927
 FT EFT 4927 4967
 FT EFT 4967 5007
 FT EFT 5007 5047
 FT EFT 5047 5087
 FT EFT 5087 5127
 FT EFT 5127 5167
 FT EFT 5167 5207
 FT EFT 5207 5247
 FT EFT 5247 5287
 FT EFT 5287 5327
 FT EFT 5327 5367
 FT EFT 5367 5407
 FT EFT 5407 5447
 FT EFT 5447 5487
 FT EFT 5487 5527
 FT EFT 5527 5567
 FT EFT 5567 5607
 FT EFT 5607 5647
 FT EFT 5647 5687
 FT EFT 5687 5727
 FT EFT 5727 5767
 FT EFT 5767 5807
 FT EFT 5807 5847
 FT EFT 5847 5887
 FT EFT 5887 5927
 FT EFT 5927 5967
 FT EFT 5967 6007
 FT EFT 6007 6047
 FT EFT 6047 6087
 FT EFT 6087 6127
 FT EFT 6127 6167
 FT EFT 6167 6207
 FT EFT 6207 6247
 FT EFT 6247 6287
 FT EFT 6287 6327
 FT EFT 6327 6367
 FT EFT 6367 6407
 FT EFT 6407 6447
 FT EFT 6447 6487
 FT EFT 6487 6527
 FT EFT 6527 6567
 FT EFT 6567 6607
 FT EFT 6607 6647
 FT EFT 6647 6687
 FT EFT 6687 6727
 FT EFT 6727 6767
 FT EFT 6767 6807
 FT EFT 6807 6847
 FT EFT 6847 6887
 FT EFT 6887 6927
 FT EFT 6927 6967
 FT EFT 6967 7007
 FT EFT 7007 7047
 FT EFT 7047 7087
 FT EFT 7087 7127
 FT EFT 7127 7167
 FT EFT 7167 7207
 FT EFT 7207 7247
 FT EFT 7247 7287
 FT EFT 7287 7327
 FT EFT 7327 7367
 FT EFT 7367 7407
 FT EFT 7407 7447
 FT EFT 7447 7487
 FT EFT 7487 7527
 FT EFT 7527 7567
 FT EFT 7567 7607
 FT EFT 7607 7647
 FT EFT 7647 7687
 FT EFT 7687 7727
 FT EFT 7727 7767
 FT EFT 7767 7807
 FT EFT 7807 7847
 FT EFT 7847 7887
 FT EFT 7887 7927
 FT EFT 7927 7967
 FT EFT 7967 8007
 FT EFT 8007 8047
 FT EFT 8047 8087
 FT EFT 8087 8127
 FT EFT 8127 8167
 FT EFT 8167 8207
 FT EFT 8207 8247
 FT EFT 8247 8287
 FT EFT 8287 8327
 FT EFT 8327 8367
 FT EFT 8367 8407
 FT EFT 8407 8447
 FT EFT 8447 8487
 FT EFT 8487 8527
 FT EFT 8527 8567
 FT EFT 8567 8607
 FT EFT 8607 8647
 FT EFT 8647 8687
 FT EFT 8687 8727
 FT EFT 8727 8767
 FT EFT 8767 8807
 FT EFT 8807 8847
 FT EFT 8847 8887
 FT EFT 8887 8927
 FT EFT 8927 8967
 FT EFT 8967 9007
 FT EFT 9007 9047
 FT EFT 9047 9087
 FT EFT 9087 9127
 FT EFT 9127 9167
 FT EFT 9167 9207
 FT EFT 9207 9247
 FT EFT 9247 9287
 FT EFT 9287 9327
 FT EFT 9327 9367
 FT EFT 9367 9407
 FT EFT 9407 9447
 FT EFT 9447 9487
 FT EFT 9487 9527
 FT EFT 9527 9567
 FT EFT 9567 9607
 FT EFT 9607 9647
 FT EFT 9647 9687
 FT EFT 9687 9727
 FT EFT 9727 9767
 FT EFT 9767 9807
 FT EFT 9807 9847
 FT EFT 9847 9887
 FT EFT 9887 9927
 FT EFT 9927 9967
 FT EFT 9967 10007
 FT EFT 10007 10047
 FT EFT 10047 10087
 FT EFT 10087 10127
 FT EFT 10127 10167
 FT EFT 10167 10207
 FT EFT 10207 10247
 FT EFT 10247 10287
 FT EFT 10287 10327
 FT EFT 10327 10367
 FT EFT 10367 10407
 FT EFT 10407 10447
 FT EFT 10447 10487
 FT EFT 10487 10527
 FT EFT 10527 10567
 FT EFT 10567 10607
 FT EFT 10607 10647
 FT EFT 10647 10687
 FT EFT 10687 10727
 FT EFT 10727 10767
 FT EFT 10767 10807
 FT EFT 10807 10847
 FT EFT 10847 10887
 FT EFT 10887 10927
 FT EFT 10927 10967
 FT EFT 10967 11007
 FT EFT 11007 11047
 FT EFT 11047 11087
 FT EFT 11087 11127
 FT EFT 11127 11167
 FT EFT 11167 11207
 FT EFT 11207 11247
 FT EFT 11247 11287
 FT EFT 11287 11327
 FT EFT 11327 11367
 FT EFT 11367 11407
 FT EFT 11407 11447
 FT EFT 11447 11487
 FT EFT 11487 11527
 FT EFT 11527 11567
 FT EFT 11567 11607
 FT EFT 11607 11647
 FT EFT 11647 11687
 FT EFT 11687 11727
 FT EFT 11727 11767
 FT EFT 11767 11807
 FT EFT 11807 11847
 FT EFT 11847 11887
 FT EFT 11887 11927
 FT EFT 11927 11967
 FT EFT 11967 12007
 FT EFT 12007 12047
 FT EFT 12047 12087
 FT EFT 12087 12127
 FT EFT 12127 12167
 FT EFT 12167 12207
 FT EFT 12207 12247
 FT EFT 12247 12287
 FT EFT 12287 12327
 FT EFT 12327 12367
 FT EFT 12367 12407
 FT EFT 12407 12447
 FT EFT 12447 12487
 FT EFT 12487 12527
 FT EFT 12527 12567
 FT EFT 12567 12607
 FT EFT 12607 12647
 FT EFT 12647 12687
 FT EFT 12687 12727
 FT EFT 12727 12767
 FT EFT 12767 12807
 FT EFT 12807 12847
 FT EFT 12847 12887
 FT EFT 12887 12927
 FT EFT 12927 12967
 FT EFT 12967 13007
 FT EFT 13007 13047
 FT EFT 13047 13087
 FT EFT 13087 13127
 FT EFT 13127 13167
 FT EFT 13167 13207
 FT EFT 13207 13247
 FT EFT 13247 13287
 FT EFT 13287 13327
 FT EFT 13327 13367
 FT EFT 13367 13407
 FT EFT 13407 13447
 FT EFT 13447 13487
 FT EFT 13487 13527
 FT EFT 13527 13567
 FT EFT 13567 13607
 FT EFT 13607 13647
 FT EFT 13647 13687
 FT EFT 13687 13727
 FT EFT 13727 13767
 FT EFT 13767 13807
 FT EFT 13807 13847
 FT EFT 13847 13887
 FT EFT 13887 13927
 FT EFT 13927 13967
 FT EFT 13967 14007
 FT EFT 14007 14047
 FT EFT 14047 14087
 FT EFT 14087 14127
 FT EFT 14127 14167
 FT EFT 14167 14207
 FT EFT 14207 14247
 FT EFT 14247 14287
 FT EFT 14287 14327
 FT EFT 14327 14367
 FT EFT 14367 14407
 FT EFT 14407 14447
 FT EFT 14447 14487
 FT EFT 14487 14527
 FT EFT 14527 14567
 FT EFT 14567 14607
 FT EFT 14607 14647
 FT EFT 14647 14687
 FT EFT 14687 14727
 FT EFT 14727 14767
 FT EFT 14767 14807
 FT EFT 14807 14847
 FT EFT 14847 14887
 FT EFT 14887 14927
 FT EFT 14927 14967
 FT EFT 14967 15007
 FT EFT 15007 15047
 FT EFT 15047 15087
 FT EFT 15087 15127
 FT EFT 15127 15167
 FT EFT 15167 15207
 FT EFT 15207 15247
 FT EFT 15247 15287
 FT EFT 15287 15327
 FT EFT 15327 15367
 FT EFT 15367 15407
 FT EFT 15407 15447
 FT EFT 15447 15487
 FT EFT 15487 15527
 FT EFT 15527 15567
 FT EFT 15567 15607
 FT EFT 15607 15647
 FT EFT 15647 15687
 FT EFT 15687 15727
 FT EFT 15727 15767
 FT EFT 15767 15807
 FT EFT 15807 15847
 FT EFT 15847 15887
 FT EFT 15887 15927
 FT EFT 15927 15967
 FT EFT 15967 16007
 FT EFT 16007 16047
 FT EFT 16047 16087
 FT EFT 16087 16127
 FT EFT 16127 16167
 FT EFT 16167 16207
 FT EFT 16207 16247
 FT EFT 16247 16287
 FT EFT 16287 16327
 FT EFT 16327 16367
 FT EFT 16367 16407
 FT EFT 16407 16447
 FT EFT 16447 16487
 FT EFT 16487 16527
 FT EFT 16527 16567
 FT EFT 16567 16607
 FT EFT 16607 16647
 FT EFT 16647 16687
 FT EFT 16687 16727
 FT EFT 16727 16767
 FT EFT 16767 16807
 FT EFT 16807 16847
 FT EFT 16847 16887
 FT EFT 16887 16927
 FT EFT 16927 16967
 FT EFT 16967 17007
 FT EFT 17007 17047
 FT EFT 17047 17087
 FT EFT 17087 17127
 FT EFT 17127 17167
 FT EFT 17167 17207
 FT EFT 17207 17247
 FT EFT 17247 17287
 FT EFT 17287 17327
 FT EFT 17327 17367
 FT EFT 17367 17407
 FT EFT 17407 17447
 FT EFT 17447 17487
 FT EFT 17487 17527
 FT EFT 17527 17567
 FT EFT 17567 17607
 FT EFT 17607 17647
 FT EFT 17647 17687
 FT EFT 17687 17727
 FT EFT 17727 17767
 FT EFT 17767 17807
 FT EFT 17807 17847
 FT EFT 17847 17887
 FT EFT 17887 17927
 FT EFT 17927 17967
 FT EFT 17967 18007
 FT EFT 18007 18047
 FT EFT 18047 18087
 FT EFT 18087 18127
 FT EFT 18127 18167
 FT EFT 18167 18207
 FT EFT 18207 18247
 FT EFT 18247 18287
 FT EFT 18287 18327
 FT EFT 18327 18367
 FT EFT 18367 18407
 FT EFT 18407 18447
 FT EFT 18447 18487
 FT EFT 18487 18527
 FT EFT 18527 18567
 FT EFT 18567 18607
 FT EFT 18607 18647
 FT EFT 18647 18687
 FT EFT 18687 18727
 FT EFT 18727 18767
 FT EFT 18767 18807
 FT EFT 18807 18847
 FT EFT 18847 18887
 FT EFT 18887 18927
 FT EFT 18927 18967
 FT EFT 18967 19007
 FT EFT 19007 19047
 FT EFT 19047 19087
 FT EFT 19087 19127
 FT EFT 19127 19167
 FT EFT 19167 19207
 FT EFT 19207 19247
 FT EFT 19247 19287
 FT EFT 19287 19327
 FT EFT 19327 19367
 FT EFT 19367 19407
 FT EFT 19407 19447
 FT EFT 19447 19487
 FT EFT 19487 19527
 FT EFT 19527 19567
 FT EFT 19567 19607
 FT EFT 19607 19647
 FT EFT 19647 19687
 FT EFT 19687 19727
 FT EFT 19727 19767
 FT EFT 19767 19807
 FT EFT 19807 19847
 FT EFT 19847 19887
 FT EFT 19887 19927
 FT EFT 19927 19967
 FT EFT 19967 20007
 FT EFT 20007 20047
 FT EFT 20047 20087
 FT EFT 20087 20127
 FT EFT 20127 20167
 FT EFT 20167 20207
 FT EFT 20207 20247
 FT EFT 20247 20287
 FT EFT 20287 20327
 FT EFT 20327 20367
 FT EFT 20367 20407
 FT EFT 20407 20447
 FT EFT 20447 20487
 FT EFT 20487 20527
 FT EFT 20527 20567
 FT EFT 20567 20607
 FT EFT 20607 20647
 FT EFT 20647 20687
 FT EFT 20687 20727
 FT EFT 20727 20767
 FT EFT 20767 20807
 FT EFT 20807 20847
 FT EFT 20847 20887
 FT EFT 20887 20927
 FT EFT 20927 20967
 FT EFT 20967 21007
 FT EFT 21007 21047
 FT EFT 21047 21087
 FT EFT 21087 21127
 FT EFT 21127 21167
 FT EFT 21167 21207
 FT EFT 21207 21247
 FT EFT 21247 21287
 FT EFT 21287 21327
 FT EFT 21327 21367
 FT EFT 21367 21407
 FT EFT 21407 21447
 FT EFT 21447 21487
 FT EFT 21487 21527
 FT EFT 21527 21567
 FT EFT 21567 21607
 FT EFT 21607 21647
 FT EFT 21647 21687
 FT EFT 21687 21727
 FT EFT 21727 21767
 FT EFT 21767 21807
 FT EFT 21807 21847
 FT EFT 21847 21887
 FT EFT 21887

FT	DISULFID	592	605	BY SIMILARITY.
FT	DISULFID	611	622	BY SIMILARITY.
FT	DISULFID	618	631	BY SIMILARITY.
FT	DISULFID	633	646	BY SIMILARITY.
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	890	890	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	956 AA;	106779 MM;	3E4C22770B6E6EEB CRC64;

Query Match	Best Local Similarity	11.1%;	Score 339;	DB 1;	Length 956;
Matches 140;	Conservative	23.28;	Pred. No. 2e-16;	Mismatches 225;	Indels 162;
				Gaps	28

QY	25	SARHHGL--LASAOPGCHGYTKIACQYGRNR-SKGVCE----	ARCPEGCKRGECV	75
Db	366	ASSNHGCGHECVNAQTALCR-----CLKGLMNPDRKTCRINTCALNKRGCCH-ECV	418	
QY	76	GNK--CRCFGY-----TGKTCSDVDNCGMKPRPCOHCRCVNTGSHGYCFLSGHMLP	128	
Db	419	NTEEHYGCRCRGVNLDPNGKTCR-YBHCADQDHCEDLCLNTEESFVCCSEGLND	477	
QY	129	D-ATGVNSTRKAMIN--COYSCDEPHEEGHQCSPSSGLRLANGRDCLIDPCASKVIC	185	
Db	478	DLKTSRADYCLLSNHGCEYSCVNTDKSPACCP-BGHLNRBDGKTCALDSCALDHC	536	
QY	186	PYN-----RR-----CVNFGSYCCKHGF	206	
Db	537	EHSVSEDSFVCCFEGYILRDDEKTCRRKDYCODVNGCHEILCVNSESYYCKLGEF	596	
QY	207	ELQYISGRDCLIDNCTMDSHTCSHNACFTMGSEFKCKCKQGY--KNGLRCSAIPEN	264	
Db	597	RLAEDGR--CRKNVCSTQGCCH--MCVNNNGSYLCRCSEGFVLADGKHCKRCTG	652	
QY	265	SVKEVLRAPGITKRIKLLAKNSMKKAKIKRYNPEPTRP-TPKVNLOPFNVE--E	320	
Db	653	PIDLYFVIDGS-----KSLGEENEFYVHFVTGIIDLSAVPKARVGLDYSQVRE	706	
QY	321	IYRGNNGSHGKKGKDEKKKEGLEDEKREK-----ALKNDIEBSLGDVFFPVNNA	374	
Db	707	FTLRPFSS-----AKEMKKAVTIMKXYKGKSMGLAKLHMEFSEFTVEGARPPSTGV	759	
QY	375	GEFGIILVQ-----RKALTSKLEH-----KDLNISVD	401	
Db	760	PRVAIVFTDGRADQDVSEWAKAKANGITMAYAVGSKALEEDELQELASPEIDKHLFYAD	819	
QY	402	CS-----FNHGICDMKQD--REDDFDMNPADRDNAIGFYMAVPALAGHKKIDIGIK	450	
Db	820	FSTMGSEISKELEKICEALEVDGSGRDSXAMV-----LPQAHOPTEPEPT	866	
QY	451	LILPLDPOPSNFCI----LFQYRLAGDKVKGKLRVFNKSNMNLAMWEKTSSEDEKMTGKI	506	
Db	867	IKIKLLSSNFAVQHRFLFEDDNLSRSQKLFHSIKSSGNPL-----ESODQCKCEMI	921	
QY	507	QLYQ 510		
Db	922	ILFQ 925		

RESULT 11
FBLL_HUMAN
STANDARD; PRT; 703 AA.
P23142; P23143; P23144; P37886; Q9UGR4;
01-NOV-1991 (Rel. 20, Created)
01-NOV-1991 (Rel. 20, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Fibulin-1 precursor.
FBLL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
MEDLINE=91100426; PubMed=2269669;

```

RA Argaves W.S., Tran H., Burgess W.H., Dickerson K.:
RT "Fibulin is an extracellular matrix and plasma glycoprotein with
RL repeated domain structure."
RN J. Cell Biol. 111:3155-3164(1990).
RP [2]
RA Argaves S.:
RP SEQUENCE FROM N.A. (ISOFORM D).
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 567-703 FROM N.A. (ISOFORM C).
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-44.
RX MEDLINE=89354537; PubMed=2527614;
RA Argaves W.S., Dickerson K., Burgess W.H., Ruoslahti E.:
RT "Fibulin, a novel protein that interacts with the fibronectin
RL receptor beta subunit cytoplasmic domain."
RN Cell 58:623-629(1989).
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: A, B, C AND D (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER ONLY IN THEIR C-
CC TERMINAL REGIONS.
CC -1- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53741; CAA37770.1; -
DR EMBL: X53742; CAA37771.1; -
DR EMBL: X53743; CAA37772.1; -
DR EMBL: U01244; AAB17099.1; -
DR EMBL: Z95331; CAB62960.1; -
DR PIR: A32826; A32826; -
DR PIR: A36346; A36346; -
DR PIR: B36346; B36346; -
DR PIR: C36346; C36346; -
DR HSSP: P35555; 1EMN.
DR MIM: 135820; -
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF01821; ANATO_3.
DR Pfam: PF00068; EGF_6.
DR SMART: SM00104; ANATO_3.
DR SMART: SM00179; EGF_Ca_7.
DR SMART: SM00001; EGF_Like_2.
DR PROSITE: PS00010; ASX_HYDROXYL_4.
DR PROSITE: PS00022; EGF_1; FALSE NEG.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; 3.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 3.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_Ca; 8.
KW Signal; Alternative splicing; Glycoprotein; Extracellular matrix;
KW Repeat; EGF-like domain; Calcium-binding.
FT SIGNAL 1 29
FT CHAIN 30 703 FIBULIN-1.
FT DOMAIN 36 76 ANAPHYLATOXIN-LIKE 1.
FT DOMAIN 77 111 ANAPHYLATOXIN-LIKE 2.
FT DOMAIN 112 144 ANAPHYLATOXIN-LIKE 3.
FT DOMAIN 176 215 EGF-LIKE 1.
FT DOMAIN 216 261 EGF-LIKE 2.
FT DOMAIN 262 307 EGF-LIKE 3.
FT DOMAIN 308 355 EGF-LIKE 4.
FT DOMAIN 356 398 EGF-LIKE 5.
FT DOMAIN 399 440 EGF-LIKE 6.

```

```

FT DOMAIN 441 480 EGF-Like 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 481 524 EGF-Like 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 525 578 EGF-Like 9, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 36 61 BY SIMILARITY.
FT DISULFID 37 68 BY SIMILARITY.
FT DISULFID 50 69 BY SIMILARITY.
FT DISULFID 78 109 BY SIMILARITY.
FT DISULFID 91 110 BY SIMILARITY.
FT DISULFID 112 136 BY SIMILARITY.
FT DISULFID 113 143 BY SIMILARITY.
FT DISULFID 126 144 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 186 199 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 220 233 BY SIMILARITY.
FT DISULFID 227 242 BY SIMILARITY.
FT DISULFID 248 260 BY SIMILARITY.
FT DISULFID 266 279 BY SIMILARITY.
FT DISULFID 273 288 BY SIMILARITY.
FT DISULFID 294 306 BY SIMILARITY.
FT DISULFID 312 325 BY SIMILARITY.
FT DISULFID 319 334 BY SIMILARITY.
FT DISULFID 341 354 BY SIMILARITY.
FT DISULFID 360 373 BY SIMILARITY.
FT DISULFID 367 382 BY SIMILARITY.
FT DISULFID 384 397 BY SIMILARITY.
FT DISULFID 403 415 BY SIMILARITY.
FT DISULFID 411 424 BY SIMILARITY.
FT DISULFID 426 439 BY SIMILARITY.
FT DISULFID 445 454 BY SIMILARITY.
FT DISULFID 450 463 BY SIMILARITY.
FT DISULFID 465 479 BY SIMILARITY.
FT DISULFID 485 498 BY SIMILARITY.
FT DISULFID 494 507 BY SIMILARITY.
FT DISULFID 509 523 BY SIMILARITY.
FT DISULFID 529 542 BY SIMILARITY.
FT DISULFID 536 551 BY SIMILARITY.
FT DISULFID 556 577 BY SIMILARITY.
FT CARBOHYD 98 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASAPLIC 567 703 MISSING (IN ISOPFORM A).
FT VASAPLIC 567 601 LQKPTDPRCTKSCRPNDVTCVDPVTHSHYI -> QK
FT SKGRONTPASGSKEDCVLWKGLEPTHLDA (IN
FT ISOFORM B).
FT VASAPLIC 567 703 LQKPTDPRCTKSCRPNDVTCVDPVTHSHYI
FT EFTREELIIFLATPPEPAPAOANIIFDITGNLSDPDI
FT KRYMDGTVGVROVRPPEPPEHATKEMNYVGVGVSHR
FT NVNVVRTPSEVME -> RCRRLCHEMRKESKIPRTTYF
FT HLSRPNIQAPAVVFRMGPSAVFEDSKQLITGNEGEFF
FT TTRKVPSPHSGVALTKPVEPERDLITVKKMLSRGTVSSF
FT VAKLFIYSAEL (IN ISOPFORM C).
FT C -> S (IN REF. 4).
FT HR -> SH (IN REF. 4).
FT CONFLICT 36 36
FT CONFLICT 41 42
FT SEQUENCE 703 AA: 77274 MW: EFD8465BA2D3A25 CXC64;
SO
Query Match 11.0%, Score 338; DB 1; Length 703;
Best Local Similarity 34.9%, Pred. No. 1.7e-16;
Matches 88; Conservative 30; Mismatches 86; Indels 48; Gaps 15;

```

```

Db 460 SYQCCRRNOYQLSDVDG-VTCEDIDECALPTGHCITCY--RCINIPGSPCCSPSSGYRL 516
QY 253 -GNGLRCSAIP 263
Db 517 APNGRNCODIDE 528

RESULT 12
NOTC_DROME
ID NOTC_DROME STANDARD: PRT: 2703 AA.
AC P07207; P04154; O97458; Q9W4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus Notch protein precursor.
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG39396.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Oregon-R; TISSUE-Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats."
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Canton-S; and Oregon-R; TISSUE-Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors."
RL Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtill J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beous P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.D., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Iel Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., Meleod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```


RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*,";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Dreano S., Gloux S., Leleau V., Mottier S., Galibert F., Borkova D.,
 RA Milana B., Kafatos F.C., Louis C., Siden-Kiamis I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Mcdolell J., Peter A., Schoettler P., Werner M., Moutkoti F.,
 RA Belnert N., Dove G., Schaefer U., Jaackie H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamsiou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*,";
 RL Science 287:2220-2222(2000).
 RN [5]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85099329; PubMed=2981631.
 RA Wharton K.A., Yedvodnick B., Flinerty V.G., Attavanis-Tsakonas S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in *D. melanogaster*,";
 RL Cell 40:55-62(1985).
 RN [6]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
 RT "Restriction of P-element insertions at the Notch locus of *Drosophila melanogaster*,";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [7]
 RP REVIEW.
 RA Harris W.A.;
 RT "Many cell types specified by Notch function,";
 RL Curr. Biol. 1:120-122(1991).
 CC -1- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M16152; AAB59220.1; -;
 DR EMBL: M16153; AAB59220.1; JOINED.
 DR EMBL: M16149; AAB59220.1; JOINED.
 DR EMBL: M16150; AAB59220.1; JOINED.
 DR EMBL: M16151; AAB59220.1; JOINED.
 DR EMBL: K03508; AAA28725.1; -;
 DR EMBL: M13689; AAA28725.1; JOINED.
 DR EMBL: K03507; AAA28725.1; JOINED.
 DR EMBL: AE003426; AAF45848.2; -;

DR EMBL: AL035436; CAB37610.1; -;
 DR EMBL: AL035395; CAB37610.1; JOINED.
 DR EMBL: M12175; AAA74496.1; -;
 DR EMBL: M16025; AAA28726.1; -;
 DR PIR: A24420; A24420.
 DR PIR: A24768; A24768.
 DR PIR: A05267; A05267.
 DR HSSP: P00740; 11XA.
 DR FLYBase: FBgn0004647; N.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF-IT.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PF00008; ank; 6.
 DR Pfam: PF00008; EGF; 36.
 DR Pfam: PF00066; notch; 3.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR01452; NOTCH.
 DR SMART: SM00248; ANK; 4.
 DR SMART: SM00179; EGF_CA; 23.
 DR SMART: SM00001; EGF-like; 13.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 34.
 DR PROSITE: PS01186; EGF-2; 28.
 DR PROSITE: PS01187; EGF_CA; 22.
 KW Differentiation: Neurogenesis; Repeat: ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 44
 FT CHAIN 45 2703
 FT DOMAIN 45 1745
 FT TRANSMM 1746 1766
 FT DOMAIN 1767 2703
 FT DOMAIN 58 95
 FT DOMAIN 96 136
 FT DOMAIN 139 176
 FT DOMAIN 177 215
 FT DOMAIN 217 253
 FT DOMAIN 255 291
 FT DOMAIN 293 329
 FT DOMAIN 331 370
 FT DOMAIN 372 408
 FT DOMAIN 409 447
 FT DOMAIN 449 486
 FT DOMAIN 488 524
 FT DOMAIN 524 562
 FT DOMAIN 564 600
 FT DOMAIN 602 637
 FT DOMAIN 639 675
 FT DOMAIN 677 713
 FT DOMAIN 715 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 865
 FT DOMAIN 867 905
 FT DOMAIN 907 944
 FT DOMAIN 946 982
 FT DOMAIN 984 1020
 FT DOMAIN 1022 1058
 Query Match 11.0%; Score 335.5; DB 1; Length 2703;
 Best Local Similarity 26.4%; Pred. No. 1,le-15;
 Matches 96; Conservative 34; Mismatches 106; Indels 127; Gaps 17;
 QY 39 GYCHYGT-----KLACCYG--RRNSKGVCEFT-CEPGCKRGEYGPKN--CRCPFGTYG 88
 Db 152 GYCHYGTLEEVYACANGYGERCTKNLCASSPCRNAGATCTALAGSSSFTCCSPPGFTG 211

Matches 86; Conservative 31; Mismatches 87; Indels 48; Gaps 15;

QY 46 KLACCYGRRRSKVC-----EATCEGCKFGEVGNKCR-----CEPGV-----T 87
Db 293 KLOKSGFIQALNCIDINECLISAPCVGQFCINTEGSYTCKQNPNGRGLHNEE 352
QY 88 KGTCSQDVNECGMKPRPC--OHRCVNTHSGYKFCFLSGHMMMPATCVNSRTCAIN--- 142
Db 353 GTRC-VVYDECAPPAEPGCKGHCLNSGSRCECKAGFYF--DGI---SRTCYDINCO 406
QY 143 -----COYSEDTBEGBQCLCPSSGLRLAPNGDCLDIDECASGKYICPNRRCVNTFG 196
Db 407 RYPRRLGCHKENTPGSFHSCS-SAGEFLSYDGRSCDEYNCLMS---PCSOECANVYG 461
QY 197 SYVCKCHIGFELQYISGRYDIDINECTMDS--HTCSHANCFTMGSEFKCKG-KQGYK- 252
Db 462 SYGTCYCRGCTSDVDG-VTCEDIDECALPTGHCIST--RCINIPGSFQCSPPSGYRL 518
QY 253 -GNGLRCSAIRE 263
Db 519 APNGRNCQDIDE 530

RESULT 14
NOTC_BRARE
ID NOTC_BRARE STANDARD; PRT; 2437 AA.
AC P46530:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein precursor.
GN NOTCH OR NOTCH1A.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993).
CC -1- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC EMBRIO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED.
CC ACCUMULATING PREDOMINANTLY IN THE PRECHORAL MESODERM AND
CC NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC BRAIN AND HEAD REGIONS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69088; GAA48831.1; -
DR HSSP: P00740; IEDM.
DR ZFIN: ZDB-GENE-990415-173; notchla.
DR InterPro: IPR002110; ANK.

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_11.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 36.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBLDOD.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 19.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 23.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 28.
DR PROSITE: PS01187; EGF_CA; 22.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
KW SIGNAL
FT CHAIN 21 2437
FT DOMAIN 21 1724
FT TRANSMEM 1725 1747
FT DOMAIN 1748 2437
FT DOMAIN 21 57
FT DOMAIN 58 98
FT DOMAIN 101 138
FT DOMAIN 139 175
FT DOMAIN 177 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 524
FT DOMAIN 526 562
FT DOMAIN 564 599
FT DOMAIN 601 637
FT DOMAIN 639 674
FT DOMAIN 676 712
FT DOMAIN 714 749
FT DOMAIN 751 787
FT DOMAIN 789 825
FT DOMAIN 827 865
FT DOMAIN 867 903
FT DOMAIN 905 941
FT DOMAIN 943 979
FT DOMAIN 981 1017
FT DOMAIN 1019 1055
FT DOMAIN 1057 1093
FT DOMAIN 1095 1141
FT DOMAIN 1143 1179
FT DOMAIN 1181 1217
FT DOMAIN 1219 1263
FT DOMAIN 1265 1303
FT DOMAIN 1305 1344
FT DOMAIN 1346 1382
FT DOMAIN 1385 1423
FT REPEAT 1446 1486
FT REPEAT 1487 1520
FT REPEAT 1521 1561
FT REPEAT 1867 1910
FT REPEAT 1915 1944
FT REPEAT 1948 1978
FT REPEAT 1982 2011
FT REPEAT 2015 2044
FT REPEAT 2048 2077
POTENTIAL.
EXTRACELLULAR (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
EGF-LIKE 1.
EGF-LIKE 2.
EGF-LIKE 3.
EGF-LIKE 4.
EGF-LIKE 5.
EGF-LIKE 6.
EGF-LIKE 7.
EGF-LIKE 8.
EGF-LIKE 9.
EGF-LIKE 10.
EGF-LIKE 11.
EGF-LIKE 12.
EGF-LIKE 13.
EGF-LIKE 14.
EGF-LIKE 15.
EGF-LIKE 16.
EGF-LIKE 17.
EGF-LIKE 18.
EGF-LIKE 19.
EGF-LIKE 20.
EGF-LIKE 21.
EGF-LIKE 22.
EGF-LIKE 23.
EGF-LIKE 24.
EGF-LIKE 25.
EGF-LIKE 26.
EGF-LIKE 27.
EGF-LIKE 28.
EGF-LIKE 29.
EGF-LIKE 30.
EGF-LIKE 31.
EGF-LIKE 32.
EGF-LIKE 33.
EGF-LIKE 34.
EGF-LIKE 35.
EGF-LIKE 36.
LIN/NOTCH 1.
LIN/NOTCH 2.
LIN/NOTCH 3.
ANK 1.
ANK 2.
ANK 3.
ANK 4.
ANK 5.
ANK 6.

```

FT DOMAIN 2265 2276 POLY-GLN (OPA-REPEAT).
FT DISULFID 25 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 86 BY SIMILARITY.
FT DISULFID 88 97 BY SIMILARITY.
FT DISULFID 105 116 BY SIMILARITY.
FT DISULFID 110 126 BY SIMILARITY.
FT DISULFID 128 137 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 181 194 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 578 BY SIMILARITY.
FT DISULFID 573 587 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.

```

```

FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
FT DISULFID 1023 1034 BY SIMILARITY.
FT DISULFID 1028 1043 BY SIMILARITY.
FT DISULFID 1045 1054 BY SIMILARITY.

Query Match 10.8%, Score 332; DB 1; Length 2437;
Best Local Similarity 28.6%, Pred. No. 1.8e-15;
Matches 95; Conservative 26; Mismatches 99; Indels 112; Gaps 17;

QY 39 GYCHT---GTLACCTG---RNSKVCBATCPCG-----CKRGE--- 73
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 GTCHNTIGFHCYCVGNWGTGDSCSEINIDCASAAASHATCHDRVASFECRPHRTGL 368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 -----CY-----GPNKCRCPGYTGTCSDVYEGMKRPPCOH---RC 109
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 369 CHLDACISNFCOKSGNCDTNPVSGKAICTCPGGTGSACNODIDECSLGANPCBHGRC 428
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 VNTGSYKFCFLSGH-----MLMPDTC-----V 133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 429 INTKSFQCKGLQGEYGRCEMDVNECKSNPCQNDATCLDQIGFHCICMPYEGVFCQI 488
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 NSRTCAMINC-QYSCEDTBEQPCLCPSGLRLAPNGRDC-LDIDECASGAYICPYNRRC 191
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 489 NSDDCASQPCNLNGKCIDINSFHCPCPKGF-----SGSLCQVDDECAS--TPCKNGAK 541
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 192 VNTFGSYCKCHIGELAYISGRYDCIDINECTMDSHFCSHHANCFNFGSPFKCKOGY 251
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 542 TIDGPKRYTCECTPGF-----SGIHCELDINECA--SSPC-HYGVCRDGVASTCGRRCGY 593
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 KNGILCSAIPENSVKEYVLAP---GTIKDR 279
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 594 TGR-----LCETINIECLASQPCRNCGTCODR 619
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
MTN2_HUMAN STANDARD; PRT; 956 AA.
AC 000339; QNSZL1;
DT 30-MAY-2000 (Rel. 39, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Matrilin-2 precursor.
GN MATN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX PubMed-11124542;
RA Muratoglu S., Krysan K., Balazs M., Sheng H., Zakany R., Modis L.,
RA Kiss I., Deak F.;
RT "Primary structure of human matrilin-2, chromosome location of the
RT MATN2 gene and conservation of an AT-AC intron in matrilin genes.";
RL Cytogenet. Cell Genet. 90:323-327(2000).
RN [2]
RP SEQUENCE OF 644-956 FROM N.A.
RX MEDLINE=97238863; PubMed=9083061;
RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;
RT "Primary structure and expression of matrilin-2, the closest relative
RT of cartilage matrix protein within the von Willebrand factor type A-
RL J. Biol. Chem. 272:9268-9274(1997).
RN [3]
RP SEQUENCE OF 244-956 FROM N.A.
RC Tissue=Testis;
RA Duesterhoeft A., Lauber J., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN MATRIX ASSEMBLY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A

```

SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- SIMILARITY: CONTAINS 2 VFMA-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U69263; AAC51260.2; -
 DR EMBL; AL137638; CAB70853.1; ALT_INIT.
 DR HSSP; P05099; 1A05.
 DR MIM; 602108; -
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR002035; VFMA.
 DR Pfam; PF00008; EGF; 8.
 DR Pfam; PF00092; vwa; 2.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00181; EGF; 10.
 DR SMART; SM00179; EGF_CA; 10.
 DR SMART; SM00327; VMA; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 9.
 DR PROSITE; PS01186; EGF_2; 9.
 DR PROSITE; PS50234; VFMA; 2.
 DR EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil;
 KW EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil;
 KW Alternative splicing.

FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 57 956 MATRILIN-2.
 FT DOMAIN 24 232 VFMA 1.
 FT DOMAIN 238 278 EGF-Like 1.
 FT DOMAIN 279 319 EGF-Like 2.
 FT DOMAIN 320 360 EGF-Like 3.
 FT DOMAIN 361 401 EGF-Like 4.
 FT DOMAIN 402 442 EGF-Like 5.
 FT DOMAIN 443 483 EGF-Like 6.
 FT DOMAIN 484 524 EGF-Like 7.
 FT DOMAIN 525 565 EGF-Like 8.
 FT DOMAIN 566 606 EGF-Like 9.
 FT DOMAIN 607 647 EGF-Like 10.
 FT DOMAIN 655 830 VFMA 2.
 FT DOMAIN 917 955 COILED COIL (POTENTIAL).
 FT DISULFID 242 253 BY SIMILARITY.
 FT DISULFID 249 262 BY SIMILARITY.
 FT DISULFID 264 277 BY SIMILARITY.
 FT DISULFID 283 294 BY SIMILARITY.
 FT DISULFID 290 303 BY SIMILARITY.
 FT DISULFID 305 318 BY SIMILARITY.
 FT DISULFID 324 335 BY SIMILARITY.
 FT DISULFID 331 344 BY SIMILARITY.
 FT DISULFID 346 359 BY SIMILARITY.
 FT DISULFID 365 376 BY SIMILARITY.
 FT DISULFID 372 385 BY SIMILARITY.
 FT DISULFID 387 400 BY SIMILARITY.
 FT DISULFID 406 417 BY SIMILARITY.
 FT DISULFID 413 426 BY SIMILARITY.
 FT DISULFID 428 441 BY SIMILARITY.
 FT DISULFID 447 458 BY SIMILARITY.
 FT DISULFID 467 482 BY SIMILARITY.
 FT DISULFID 488 499 BY SIMILARITY.
 FT DISULFID 495 508 BY SIMILARITY.
 FT DISULFID 510 523 BY SIMILARITY.
 FT DISULFID 529 540 BY SIMILARITY.
 FT DISULFID 536 549 BY SIMILARITY.
 FT DISULFID 551 564 BY SIMILARITY.
 FT DISULFID 570 581 BY SIMILARITY.
 FT DISULFID 577 590 BY SIMILARITY.
 FT DISULFID 592 605 BY SIMILARITY.

FT DISULFID 611 622 BY SIMILARITY.
 FT DISULFID 618 631 BY SIMILARITY.
 FT DISULFID 633 646 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 890 890 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT VARSPLIC 861 879 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 356 356 K -> E (IN REF. 3).
 FT CONFLICT 594 594 V -> E (IN REF. 3).
 FT CONFLICT 644 644 R -> G (IN REF. 2).
 FT CONFLICT 755 755 F -> L (IN REF. 3).
 FT CONFLICT 935 935 L -> F (IN REF. 2).
 SQ SEQUENCE 956 AA; 106840 MW; 826B7F347178FC80 CRC64;

Query Match 10.8%; Score 331; DB 1; Length 956;
 Best Local Similarity 23.8%; Pred. No. 7.3e-16;
 Matches 129; Conservative 64; Mismatches 206; Indels 144; Gaps 24;

QY 41 CHYGTKLACCYGWRNR-SKVCE-----ATCEPGKFGCEVGPKN---CRCEPGYT----- 87
 DB 385 CH-----CLKGFTLNPDKTKCRINYNALNKPGEH-ECVMMESYCRCHRGYTLDPN 437
 QY 88 GKTCSQDVNECGMKPRPCOHRCVNTHGSKYKFCFLSGHMLMPD-ATCVNSRTCAMIN--CQ 144
 DB 438 GKTCSR-VDHCAQODHGEFOLCLNTEDSEFVCCSGEFLINEDLTKCSRYDYCLSDHGE 496
 QY 145 YSCEDTEEGPQCLCPSSGRLAPNRDCLDIDECASGVKICYNN----- 188
 DB 497 YSCVNMDSFACQCP-EGHVLNSDKGTAKLSDSCALGDHGCSCVSSSEDSFVCCFEGY 555
 QY 189 -----RR-----CVMTFGSYCKCHIGFELQYISGRYCIDINECTM 225
 DB 556 ILREDGKTCRRRDVCGALDHGEHCIVNSDSDYTECLVGFILADGKR--CRRDVCYS 613
 QY 226 DSHTSHNANCPNTOGSEKCKQGY--KGNGLRCSALPENSVKYLRAPGTIKRIKL 283
 DB 614 THNGEH--ICVNNNSYCKCSEGFVLAEDRCKCKTEGPIDLVFYIDGS-----KS 665
 QY 284 LAHKSMMKKAKIKNVTPPEPTRP--TPKVNLOPFN---YEELVSRGNSHGKKGNEEM 339
 DB 666 LGEENFEVVKOPVYTIIDSLTISPKARVGLLYSTQVHTEFTLNPN-----AKDM 718
 QY 340 KEGLEDEKREEK-----ALKNDIERSLRGDVFFPKYNEAGEFGLILVORKA----- 386
 DB 719 KKAVAHMYMGKSMTGALTKHMFERSFLOGGARPFSTRVRAAIVFDGRAQDDVSPW 778
 QY 387 -----LNSKLEHKLNTSVDCS-----FNHGICDMK 412
 DB 779 ASKAKANGITMYAVGVGAIEELOEIASEPNTKHLFYAEDPSTNDEISEKLKICICAL 838
 QY 413 QDREDDFMPNPDARDNAIGFYMAVPALAGHKHDKIRLKLDPDPOSFCLLPYRLAG 472
 DB 839 ESDSGRQD-SPAGE-----LPKTYQQPISEPVITINIDDLSCSNFAYQHRILEE 888
 QY 473 DKV 475
 DB 889 DNL 891

Search completed: October 17, 2002, 16:00:52
 Job time : 27 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 15:54:18 : Search time 31 Seconds

(without alignments)
3086.006 Million cell updates/sec

Title: US-09-687-860-24

Perfect score: 3060

Sequence: 1 MRLPWSLALPLLSWVAGF.....VDGVLVSGLCPSDLSVDD 553

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	3055	99.8	553	4	Q9NZL7	Q9NZL7 homo sapien
2	3049.5	99.7	554	4	Q9NT67	Q9NT67 homo sapien
3	3016	98.6	558	4	Q9UPE6	Q9UPE6 homo sapien
4	2401.5	78.5	550	11	Q9JF25	Q9JF25 mus musculu
5	1095	35.8	561	11	Q9JY88	Q9JY88 mus musculu
6	1081.5	35.3	578	11	Q9JY88	Q9JY88 mus musculu
7	1069.5	35.0	592	11	Q9JY88	Q9JY88 mus musculu
8	1056	34.5	592	11	Q9JY88	Q9JY88 mus musculu
9	408.5	13.3	2906	11	Q9JY88	Q9JY88 mus musculu
10	405	13.2	2809	4	Q96JF8	Q96JF8 homo sapien
11	386	12.6	708	13	P87363	P87363 gallus gall
12	385.5	12.6	2872	11	Q9WU88	Q9WU88 ratu mus norv
13	383.5	12.5	3857	11	Q9WU88	Q9WU88 mus musculu
14	382	12.5	1174	11	Q99K58	Q99K58 mus musculu
15	374.5	12.2	2189	5	Q9B105	Q9B105 eimeria ten
16	370	12.1	608	11	Q9DBE2	Q9DBE2 mus musculu

17	369	12.1	937	5	Q9BLJ1	Q9BLJ1 ciona intes
18	365.5	11.9	1574	11	Q88281	Q88281 ratu mus norv
19	362.5	11.8	576	4	Q9Y3V7	Q9Y3V7 homo sapien
20	362.5	11.8	1511	4	Q75412	Q75412 homo sapien
21	361.5	11.8	1587	4	Q00508	Q00508 homo sapien
22	354.5	11.6	955	4	Q96DN2	Q96DN2 homo sapien
23	354.5	11.6	1833	11	Q08999	Q08999 mus musculu
24	352	11.5	704	13	Q73774	Q73774 gallus gall
25	349	11.4	2673	4	Q96SC3	Q96SC3 homo sapien
26	348	11.4	685	11	Q922K8	Q922K8 mus musculu
27	346	11.3	495	4	Q9HBQ5	Q9HBQ5 homo sapien
28	340.5	11.1	1664	5	Q9TVQ2	Q9TVQ2 caenorhabdi
29	339.5	11.1	741	4	Q96K89	Q96K89 homo sapien
30	339	11.1	956	11	Q99K64	Q99K64 mus musculu
31	339	11.1	1764	11	Q35806	Q35806 ratu mus norv
32	338	11.0	528	11	Q9CXD8	Q9CXD8 mus musculu
33	338	11.0	5636	4	Q96RW7	Q96RW7 homo sapien
34	334	10.9	517	4	Q9NP01	Q9NP01 homo sapien
35	333	10.9	2447	13	Q13149	Q13149 fuigu rubrip
36	333	10.9	2653	5	Q25253	Q25253 lucilia cup
37	332	10.8	999	4	Q9N036	Q9N036 homo sapien
38	331	10.8	589	5	Q9TZS1	Q9TZS1 caenorhabdi
39	331	10.8	689	5	Q95N23	Q95N23 caenorhabdi
40	330.5	10.8	961	11	Q9EQC6	Q9EQC6 mus musculu
41	330	10.8	798	5	Q18026	Q18026 caenorhabdi
42	329.5	10.8	681	13	Q42182	Q42182 brachydanio
43	329.5	10.8	2319	11	Q9R172	Q9R172 ratu mus norv
44	328	10.7	1062	11	Q60789	Q60789 mus musculu
45	328	10.7	1713	11	Q88349	Q88349 mus musculu

ALIGNMENTS

RESULT 1
Q9NZL7 PRELIMINARY; PRT; 553 AA.
ID Q9NZL7
AC Q9NZL7
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE EPIDERMAL GROWTH FACTOR REPEAT CONTAINING PROTEIN.
GN EGF6.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20079166; PubMed=10610727;
RA Yeung G., Mulero J.J., Berntsen R.P., Loeb D.B., Drmanac R.,
Ford J.E.;
RT "Cloning of a novel epidermal growth factor repeat containing gene
RT EGF6: expressed in tumor and fetal tissues.";
RL Genomics 62:304-307(1999).
CC -1 SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF186084; AAF27812.1; -.
DR HSBP; P00736; IAPQ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF_Like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000998; MAM.
DR Pfam: PF00629; MAM; 1.
DR SMART: SM00179; EGF_CA; 3.
DR SMART: SM00001; EGF_Like; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 3.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 3.
DR PROSITE: PS00060; MAM_2; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 553 AA; 61314 MW; 2F555F167857DE50 CRC64;

Query Match 99.8%; Score 3055; DB 4; Length 553;
 Best Local Similarity 99.8%; Pred. No. 2,1e-240;
 Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFLPWSLALPLLSSVAGFGMAASARRHHGLASARQPGVCHYGTKLACCYGWRNSKGV 60
 DB 1 MFLPWSLALPLLSSVAGFGMAASARRHHGLASARQPGVCHYGTKLACCYGWRNSKGV 60
 QY 61 CEATCEPGCKGEGECVGPKNCRCPFGYTGTCSQDVNECGMKRPPQOHCVNTHGSKFC 120
 DB 61 CEATCEPGCKGEGECVGPKNCRCPFGYTGTCSQDVNECGMKRPPQOHCVNTHGSKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCQYSCDETEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCQYSCDETEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
 QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 DB 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 QY 241 GSFCKCKOGYKGNGLRCSAIPENSVKEVLRAPGTIKDRIRKLLAHKNSMKKAKIKNTV 300
 DB 241 GSFCKCKOGYKGNGLRCSAIPENSVKEVLRAPGTIKDRIRKLLAHKNSMKKAKIKNTV 300
 QY 301 PEPTPTPKVNLQPFNYEEIYSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
 DB 301 PEPTPTPKVNLQPFNYEEIYSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
 QY 361 SLRGVFFPKVNEGEFGLILVQRKALTSKLEHKDLNLSVDCSFNHGICDMKODREDDF 420
 DB 361 SLRGVFFPKVNEGEFGLILVQRKALTSKLEHKDLNLSVDCSFNHGICDMKODREDDF 420
 QY 421 WNPADRONAIGFYMAVAPALAGHKIDIGRLKLLPDLQPOSNCLLPDYRLAGDVKGLRV 480
 DB 421 WNPADRONAIGFYMAVAPALAGHKIDIGRLKLLPDLQPOSNCLLPDYRLAGDVKGLRV 480
 QY 481 FVKNSNNALAMEKTTSEDEKWKTKIQLYQGTDAKTSIIFEAERKKGTEIAVDGVLLV 540
 DB 481 FVKNSNNALAMEKTTSEDEKWKTKIQLYQGTDAKTSIIFEAERKKGTEIAVDGVLLV 540
 QY 541 SGLCPDLSLSD 553
 DB 541 SGLCPDLSLSD 553

RESULT 2
 Q9NY67 PRELIMINARY; PRT: 554 AA.

AC Q9NY67; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHEtical 61.4 KDA PROTEIN.
 GN W80.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=TERATOCARCINOMA/NEURON;
 RA Franco B.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TERATOCARCINOMA/NEURON;
 RX MEDLINE=20241927; PubMed=10777661;
 RA Buchner G., Orfanelli U., Quaderl N., Bassi M.T., Andolfi G.;
 RT Identification of a new EGF-repeat-containing gene from human Xp22:
 RT Acanditate for developmental disorders.";
 RL Genomics 65:16-23(2000).
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

DR EMBL; AJ245671; CAB92132.1; -.
 DR HSSP; P00736; IAPQ.
 DR InterPro; IPR000152; Asx_Hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR000998; MAM.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00629; MAM; 1.
 DR SMART; SM00179; EGF_CA; 3.
 DR SMART; SM00001; EGF_Like; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR PROSITE; PS00060; MAM_2; 1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
 KW Hypothetical protein; Repeat.
 SEQUENCE 554 AA: 61388 MW: D519238F2A604101 CRC64;

Query Match 99.7%; Score 3049.5; DB 4; Length 554;
 Best Local Similarity 99.8%; Pred. No. 6e-240;
 Matches 553; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MFLPWSLALPLLSSVAGFGMAASARRHHGLASARQPGVCHYGTKLACCYGWRNSKGV 60
 DB 1 MFLPWSLALPLLSSVAGFGMAASARRHHGLASARQPGVCHYGTKLACCYGWRNSKGV 60
 QY 61 CEATCEPGCKGEGECVGPKNCRCPFGYTGTCSQDVNECGMKRPPQOHCVNTHGSKFC 120
 DB 61 CEATCEPGCKGEGECVGPKNCRCPFGYTGTCSQDVNECGMKRPPQOHCVNTHGSKFC 120
 QY 121 LSGHMLMPDATCVNSRTCAMINCQYSCDETEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
 DB 121 LSGHMLMPDATCVNSRTCAMINCQYSCDETEBGPCLCPSSGLRLAPNGRCLDIDECAS 180
 QY 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 DB 181 GKVICPYNRRCVNTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSHHANCFNTQ 240
 QY 241 GSFCKCKOGYKGNGLRCSAIPENSVKEVLRAPGTIKDRIRKLLAHKNSMKKAKIKNTV 300
 DB 241 GSFCKCKOGYKGNGLRCSAIPENSVKEVLRAPGTIKDRIRKLLAHKNSMKKAKIKNTV 300
 QY 301 PEPTPTPKVNLQPFNYEEIYSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
 DB 301 PEPTPTPKVNLQPFNYEEIYSRGNSHGKKGNEEKMEKGLDEKREKALKNDIEER 360
 QY 361 SLRGVFFPKVNEGEFGLILVQRKALTSKLEHK-DLWISVDCSFNHGICDMKODREDDF 419
 DB 361 SLRGVFFPKVNEGEFGLILVQRKALTSKLEHKDLNLSVDCSFNHGICDMKODREDDF 420
 QY 420 WNPADRONAIGFYMAVAPALAGHKIDIGRLKLLPDLQPOSNCLLPDYRLAGDVKGLRV 479
 DB 420 WNPADRONAIGFYMAVAPALAGHKIDIGRLKLLPDLQPOSNCLLPDYRLAGDVKGLRV 480
 QY 481 FVKNSNNALAMEKTTSEDEKWKTKIQLYQGTDAKTSIIFEAERKKGTEIAVDGVLL 540
 DB 481 FVKNSNNALAMEKTTSEDEKWKTKIQLYQGTDAKTSIIFEAERKKGTEIAVDGVLL 540
 QY 540 VSGLCPSLSD 553
 DB 541 VSGLCPSLSD 554

RESULT 3
 Q9UFG6 PRELIMINARY; PRT: 558 AA.

AC Q9UFG6; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHEtical 61.8 KDA PROTEIN (FRAGMENT).

Fri Oct 18 11:29:56 2002

us-09-687-860-24.rsp

```
GN DKF2P564P2063.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN
[1]
SEQUENCE FROM N.A.
RP TISSUE=BRIN;
RA Diesterheft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AL117610; CAB56014.1; -.
DR HSSP; P00736; IAPQ.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00008; EGF; 4.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS50060; MAM_2; 1.
KW Calcium-binding; EGF-like domain; glycoprotein; Hydroxylation;
KW Hypothetical protein; Repeat.
FT NON_TER 1
SQ SEQUENCE 558 AA; 61828 MW; AA38D7DCE402BFA3 CRC64;

Query Match 98.6%; Score 3016; DB 4; Length 558;
Best Local Similarity 98.9%; Pred. No. 3.2e-237;
Matches 547; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRLPMSIALPRLLSVWAGFGNNAASARHGHGLASAROPGVCHYGTIKLACCYGWRNSKGV 60
DB 6 MRLPMSIALPRLLSVWAGFGNNAASARHGHGLASAROPGVCHYGTIKLACCYGWRNSKGV 65
QY 61 CEATCEPGCKGEGCVGNPKRCRCFPGYTGKTSQDVNECGMKPRPCQHRVCVTHGSYKFC 120
DB 66 CEATCEPGCKGEGCVGNPKRCRCFPGYTGKTSQDVNECGMKPRPCQHRVCVTHGSYKFC 125
QY 121 LSGHMLPDAVCVNSRCAAMINCOYSCEDTEEGPQCICPSSGLRLAPRGRCLOIDECAS 180
DB 126 LSGHMLPDAVCVNSRCAAMINCOYSCEDTEEGPQCICPSSGLRLAPRGRCLOIDECAS 185
QY 181 GKVICPNRRRCVNTFGSYCCKHIGFELQYISGRYDCIDINECTMDSHSCSHANCFNTQ 240
DB 186 GKVICPNRRRCVNTFGSYCCKHIGFELQYISGRYDCIDINECTMDSHSCSHANCFNTQ 245
QY 241 GSFCKCKGKGYKGLRCSAIPENSVEVLAPGTIKRIKLLAHKNSMKKKAKIKVNT 300
DB 246 GSFCKCKGKGYKGLRCSAIPENSVEVLAPGTIKRIKLLAHKNSMKKKAKIKVNT 305
QY 301 PEPRTRTPKKNLQPPNVEELVSRGNSHGKKGNEEMKKGLEDEKKEEKALKNDIEER 360
DB 306 PEPRTRTPKKNLQPPNVEELVSRGNSHGKKGNEEMKKGLEDEKKEEKALKNDIEER 365
QY 361 SLRGDVEFPKYNAGEFGLILVORKALTSKLEHNDLNTSVDCSFNHGICDMKODREDFD 420
DB 366 SLRGDVEFPKYNAGEFGLILVORKALTSKLEHNDLNTSVDCSFNHGICDMKODREDFD 425
QY 421 WNPADRONAIGFYMAVRLAGHKDIGRLKLLPDLQPSNFCLLFYRLAGDKVYKRLV 480
DB 426 WNPADRONAIGFYMAVRLAGHKDIGRLKLLPDLQPSNFCLLFYRLAGDKVYKRLV 485
QY 481 FVKSNNALAMEKTTSEDEKKTGKIOLYOGTDATKSIIFPAERGKGTGELAVDGVLLV 540
DB 486 FVKSNNALAMEKTTSEDEKKTGKIOLYOGTDATKSIIFPAERGKGTGELAVDGVLLV 545
QY 541 SGLCPDLSLVVD 553
```

```
DB 546 SGLCPDLSLVVD 558
|||||
RESULT 4
ID 09JUZ5 PRELIMINARY; PRT; 550 AA.
AC 09JUZ5;
DT 01-OCT-2000 (TReMBLrel. 15 Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 61.5 KDA PROTEIN.
GN EGF6 OR W80.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
[1]
SEQUENCE FROM N.A.
RP Franco B.;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN
[2]
SEQUENCE FROM N.A.
RP MEDLINE=20241927; PubMed=10777661;
RA Buchner G., Orfanelli U., Quaderl N., Bassi M.T., Andolfi G.;
RT "Identification of a new EGF-repeat-containing gene from human Xp22:
RT Acandidate for developmental disorders.";
RL Genomics 65:16-23(2000).
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AJ245672; CAB92128.1; -.
DR HSSP; P35555; IEMN.
DR MGD; MGI:1858599; Egfl6.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000998; MAM.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS50060; MAM_2; 1.
KW Calcium-binding; EGF-like domain; glycoprotein; Hydroxylation;
KW Hypothetical protein; Repeat.
SQ SEQUENCE 550 AA; 61520 MW; DEF936325C9F31B3 CRC64;

Query Match 78.5%; Score 2401.5; DB 11; Length 550;
Best Local Similarity 77.9%; Pred. No. 3.4e-187;
Matches 431; Conservative 48; Mismatches 69; Indels 5; Gaps 4;

QY 1 MRLPMSIALPRLLSVWAGFGNNAASARHGHGLASAROPGVCHYGTIKLACCYGWRNSKGV 60
DB 1 MRLPMSIALPRLLSVWAGFGNNAASARHGHGLASAROPGVCHYGTIKLACCYGWRNSKGV 65
QY 61 CEATCEPGCKGEGCVGNPKRCRCFPGYTGKTSQDVNECGMKPRPCQHRVCVTHGSYKFC 120
DB 61 CEATCEPGCKGEGCVGNPKRCRCFPGYTGKTSQDVNECGMKPRPCQHRVCVTHGSYKFC 120
QY 121 LSGHMLPDAVCVNSRCAAMINCOYSCEDTEEGPQCICPSSGLRLAPRGRCLOIDECAS 180
DB 121 LSGHMLPDAVCVNSRCAAMINCOYSCEDTEEGPQCICPSSGLRLAPRGRCLOIDECAS 180
QY 181 GKVICPNRRRCVNTFGSYCCKHIGFELQYISGRYDCIDINECTMDSHSCSHANCFNTQ 240
DB 181 GKVICPNRRRCVNTFGSYCCKHIGFELQYISGRYDCIDINECTMDSHSCSHANCFNTQ 240
QY 241 GSFCKCKGKGYKGLRCSAIPENSVEVLAPGTIKRIKLLAHKNSMKKKAKIKVNT 300
DB 241 GSFCKCKGKGYKGLRCSAIPENSVEVLAPGTIKRIKLLAHKNSMKKKAKIKVNT 300
QY 301 PEPRTRTPKKNLQPPNVEELVSRGNSHGKKGNEEMKKGLEDEKKEEKALKNDIEER 360
DB 301 PEPRTRTPKKNLQPPNVEELVSRGNSHGKKGNEEMKKGLEDEKKEEKALKNDIEER 360
QY 361 SLRGDVEFPKYNAGEFGLILVORKALTSKLEHNDLNTSVDCSFNHGICDMKODREDFD 420
DB 361 SLRGDVEFPKYNAGEFGLILVORKALTSKLEHNDLNTSVDCSFNHGICDMKODREDFD 420
QY 421 WNPADRONAIGFYMAVRLAGHKDIGRLKLLPDLQPSNFCLLFYRLAGDKVYKRLV 480
DB 421 WNPADRONAIGFYMAVRLAGHKDIGRLKLLPDLQPSNFCLLFYRLAGDKVYKRLV 480
QY 481 FVKSNNALAMEKTTSEDEKKTGKIOLYOGTDATKSIIFPAERGKGTGELAVDGVLLV 540
DB 481 FVKSNNALAMEKTTSEDEKKTGKIOLYOGTDATKSIIFPAERGKGTGELAVDGVLLV 540
QY 541 SGLCPDLSLVVD 553
DB 541 SGLCPDLSLVVD 553
```

Oy	301	PEPPTPTPKYNLODPNPNNEEIVSRGSGNSHGKKNKEEMKGEJDEDERREKALKNIE-E	359
Db	299	PRPASTRYPKYNL-PYSEEBVSGRNTDGHQKKKEGKEERLEFEER-GEKTLRNEYDE	356
Oy	360	RSLRGDEVFEPKVNAEGEGLLYVRKALTSKLEHKDLNISVDCSFNNGIDMKWQDREDDF	419
Db	357	RTLRLDVPSPKVNAEADIEDLYVYORKETLNSKLKHKDLNISVDCSFIDGVDWKQDREDDF	416
Oy	420	DWNPADRONALGFMVAVPALAGHKHDIGRLKLLPDLQPSNCLFEDYRLADGKGLR	479
Db	417	DWNPADRONDGYTMAVYPALAGHKKNIGRLKLLPDLTPOSNCLFEDYRLADGKGLR	476
Oy	480	VFKVNSNNALAMEKTSIDEKWKMTGKIQLYQGTDATEKSIIFEARERGKGATGEIAYDGVLL	539
Db	477	VFKVNSNNALAMEETKKNEDGEMRTGKIQLYQGTDATEKSIIFEARERGKGATGEIAYDGVLL	536
Oy	540	VSGLCPPDLSL.SVD	552
Db	537	VSGLCPPDDFLSVE	549

RESULT 5	
Q91V88	
ID Q91V88	PRELIMINARY;
8182	PRT; 561 AA

DT 01-DEC-2001 (TREMBLrel_19, Created)
DT 01-DEC-2001 (TREMBLrel_19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel_19, Last annotation update)
DE POEM (NEPHROPECTIN SHORT ISOFORM).
GN NEPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA MEDLINE=21551216; PubMed=11456798;
RM Motomura N., Tezuka Y., Matsunabe N., Yasuda M., Miyatani S.,
Hozumi N., Tezuka K.,
RT "Molecular cloning of POEM, A novel adhesion molecule that interacts
RT with alpha8beta1 Integrin.";
RL J. Biol. Chem. 276:42172-42181(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS; TISSUE=KIDNEY;
RA MEDLINE=21363579; PubMed=11470831;
RM Brandeburger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
Muller U., Reichardt L.F.,
RT "Identification and characterization of a novel extracellular matrix
RT protein nephronectin that is associated with integrin alpha8beta1 in
RT the embryonic kidney.";
RL J. Cell Biol. 154:447-458(2001).
DR EMBL; AB059656; BAB96682.1; -;
DR EMBL; AY035898; AAK96010.1; -;
SQ SEQUENCE 561 AA; 61490 MW; 69E7ACAA0EE3F506 CRC64;

Query Match	35.8%	Score 1095	DB 11	Length 561
Best Local Similarity	38.0%	Pred. No. 8e-81		
Matches 210; Conservative	92;	Mismatches 180;	Indels 70;	Gaps 11;

[illegible][illegible]

RESULT 6	
Q91ZD3	
ID Q91ZD3	PRELIMINARY; PRT; 578 AA.

NC 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE NEPHRONECTIN LONG ISOFORM.
 GN NEPH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NH SWISS; TISSUE=KIDNEY;
 RX MEDLINE=21363579; PubMed=11470831;
 RA Brandenberger R., Schmidt A., Linton J., Wang D., Backus C., Denda S.,
 RA Muller U., Reichardt L.F.;
 RT "Identification and characterization of a novel extracellular matrix
 RT protein nephronectin that is associated with integrin alpha6beta1 in
 RT the embryonic kidney";
 RL J. Cell Biol. 154:447-458(2001).
 SO EMBL; AY035899; AAK96011.1; -
 SQ SEQUENCE 578 AA; 63606 MW; 4757DE54CC47DA4C CRC64;

Query Match	35.3%	Score 1081.5	DB 11	Length 578
Best Local Similarity	37.1%	Pred. No. 1e-79		
Matches 211, Conservative	92	Mismatches 179	Indels 87	Gaps 12

[illegible]

QY	39	GVHHGTRFLACCYAMRRNRKSKVCATGEPGCKREPCYGNPKRCREPEYTKTQSO-----	93
Db	34	GLCRITGGRIIDCGMAROSKSGCOPVCCPQCKHCECGPNKKCKHPGAFAGTKQODESFH	93
QY	94	-----DVNECGMKPRRCOHRNVNTHGSKYCFCLSGHMLM	127
Db	94	PTPLDQSGEQLPQPPHQANVPSRQDNEBGGLKPRCKRHKNMTFESYCYCLNGTML	153
QY	128	PDATCVNSRTCAAMINCOYSCDETEEGPOCLPSSGLRLAPNRDCLDIDECASGATICY	187
Db	154	PDGSSSALSCSMANCOYGCYVWKGQYRCQGPSGLDAPGRCIVDIDECATGRVSCPR	213
QY	188	NRCVYNTGSGYYCKCHIGFELQYISGRICLDIDIECTMDSTGCHHANCFTQGSFCKC	247
Db	214	FRQCVNTFGSYTCKCHGFDLMTYGKKQXCHDIDEGCSIGHQGSSYARCVNIHSGYCCOC	273

[illegible]

QY 278 -----DRIKLLAHKNSMKKAK-1KNVTPERTPTPKVNLQPFVVEIV 322
DB 334 RIPDAGSTRMPLKPTIPPIVITNRTSKPTTRPNPPTPPPPPLPTPTP----- 386
QY 323 SRGNSHGCKGKNEKMEGLEDKEREKAL-----KNDIERSLRGDVF 367
DB 387 -----RTTLPPTPTPTPTPTTTPATSTTTRVITVYDNRLOTDPQKPRGDVF 434
QY 368 FPKVNEAGEFGLIVORALKSLKHKDLNTSY--DCSFNHGICDMKORREDFDNPPADR 426
DB 435 IPRQPTNLFETFEIERKVSADDEYKDDPGILHSCNFDHGLCGVIREKSDLHMETAR 493
QY 427 DNATGFAVAVPALAGHKRDKIGRLKLLPDLQPSNFCLLFDYRLADGVYKLVFVK--N 484
DB 494 DPAGGQYLTVSAAKAPGKARLVRLGLHMHSGDLCSFRHKRYGLHSGTLQVFPVRKNG 553
QY 485 SNNALAMKTSSEDEKMTGKIQLYQGTDAKSTIFEAEKRGKGTGELAVDGVLLVSGIC 544
DB 554 THGALMGRNGHG--WRQTOITL-RGAD-VKSVTFKGEKRGHGTGLDVSLSKRGRC 609

RESULT 9

ID 09WU9 PRELIMINARY; PRT; 2906 AA.
AC 09WU9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIBRILLIN-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93350231; PubMed=10419698;
RA Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,
Kawar Y.S.;
RT "Cloning of rat fibrillin-2 cDNA and its role in branching
morphogenesis of embryonic lung.";
RL Dev. Biol. 212:229-242(1999).
DR EMBL; AF135060; AAC34439.1; .
DR HSSP; P35555; 1EMN.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001436; EGF_T1.
DR InterPro; IPR002212; TB.
DR Pfam; PF000683; TB; 9.
DR PRINTS; PRO0010; EGFBLD.
DR SMART; SM00179; EGF_CA_42.
DR SMART; SM00001; EGF_like; 4.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 36.
DR PROSITE; PS01187; EGF_CA; 43.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2906 AA; 313371 MW; 9EB64E72704AE58 CRC64;

Query Match 13.3%; Score 408.5; DB 11; Length 2906;
Best Local Similarity 29.3%; Pred. No. 5, 1e-24;
Matches 98; Conservative 32; Mismatches 92; Indels 113; Gaps 13;

QY 38 PGVCHYGT-----KLACCTG-----RNSKVCATCEPG--CKFGEV--GPNKC 80
DB 1073 PGMCTYGCRTMTGSGFKRCNGSFALDMERNCTDIDECRISPDLCNGMCIVNTPGSFEC 1132
QY 81 RCFPGYTG-----KTCSDVAVNCGMKPRPCQ-HRCVNTHGSGKCCSLGSHMLMP----- 128

DB 1133 ECFEGESGFMMKNC-MDIDECERNPLLCRGTCVNTGSGFQCDPLGHELSPSREDCI 1191
QY 129 -----DATVNSRTCAMI-----NCOYSCD 149
DB 1192 DINECSLDNLCNRKNCVMIMIGTYOCSNPGYQATPPDRGSCDIDECMIMNCGCTGCTN 1251
QY 150 TEEGPOCLCPSSGLRLAPNGRCDLIDECASGVIC-----PYNRCV----- 192
DB 1252 SEGSEYECSC-SEGVALMPDGRSCADIDECENNPDIJDGCGCTNIPGEYRCLCYDGFMSM 1310
QY 193 -----NTFGSYCKCHIGFELYISGRYDCIDINECTMOSH 228
DB 1311 DKMTCIDVNECDLNPNICFGECECTKGSFICHQGLYSVK--KATYGCTDVECEIGAH 1368
QY 229 TCSHANCENFQSGFKCKGKNGSLRCSAIP 263
DB 1369 NCDMHASCLNVPSFKSCSGREGVNGNGIKCIDLDE 1403

RESULT 10

ID 096JP8 PRELIMINARY; PRT; 2809 AA.
AC 096JP8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FIBRILLIN3.
GN KIAA1776.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL; AB053450; BAB47408.1; .
SQ SEQUENCE 2809 AA; 300323 MW; 20C04CC006C016F CRC64;

Query Match 13.2%; Score 405; DB 4; Length 2809;
Best Local Similarity 33.2%; Pred. No. 9, 3e-24;
Matches 93; Conservative 37; Mismatches 104; Indels 46; Gaps 12;

QY 35 ARQGVCHYGT-----KLACCTGMRNSKG-----VCEATCEPG--CKFGEV--GP 77
DB 1076 ARDPLLCRGCTNTNDGYSKCCPGHEHLLTAKGTACEDIDECSLSDGLCPHQCVAVTGA 1135
QY 78 NRCRCPGYTGKTSQ--DVNECGMKPRPCQHRVNTGSKYKCFCLSGHMLMPDATCVN 134
DB 1136 FQCSCHAGFQSTPDRQGCYVDINECRVQNGCDVHCINTGSTRCSGGQYSLMPD----- 1190
QY 135 SRTCAMIN-----C-QYSCDETEEGPOCLCPSSGRLAPNGRCDLIDECASGVIC 185
DB 1191 GRACADVDCEENPRVCDGSHCTNMPGHRCLC-YDGFATPDMRCVVDDECDLNPHTC 1249
QY 186 PYNRCVNTFGSYCKCHIGFELYISGRYDCIDINECTMDSHTSHANCCNTGSGFK 245
DB 1250 LHG-DCENTKGSFVCHQGLGYMVR--KGATGCSVDDECEVGGHNCDSHASCINTPGSF 1306
QY 246 KCKGQYKNGSLRCSAIPENSVE-----VLRAPGTK 277
DB 1307 RCLPGWVGDFECHDLDECISOEHRCSPRGDCLNVPGSTR 1346

RESULT 11
ID P87363 PRELIMINARY; PRT; 708 AA.
AC P87363;

DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FIBRILIN-1 (FRAGMENT).
 GN FBN1.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 NCBI_TaxID=9031;
 RN NCBI_TaxID=9031;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20152896; PubMed=10691037;
 RA Zhou G., Price C.E., Rosenquist T.H., Gadsen P.F., Godfrey M.,
 RT "Partial cloning and sequencing of chick fibrillin-1 cDNA."
 RL In Vitro Cell. Dev. Biol. Anim. 36:19-25(2000).
 DR EMBL; U88872; AAB48531.1; .
 DR HSSP; P07204; 2ADX.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR InterPro: IPR001491; Thrbomodulin.
 DR Pfam; PF00683; TB; 2.
 DR PRINTS; SM00907; THRBOMODULN.
 DR SMART; SM00179; EGF_CA; 14.
 DR PROSITE; PS00010; ASX_HYDROXYL; 13.
 DR PROSITE; PS01186; EGF_2; 10.
 DR PROSITE; PS01187; EGF_CA; 13.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 FT NON_TER
 SQ SEQUENCE 708 AA; 76164 MW; C247271C1DF73361 CRC64;

Query Match 12.6%; Score 386; DB 13; Length 708;
 Best Local Similarity 33.0%; Pred. No. 5.7e-23;
 Matches 89; Conservative 33; Mismatches 88; Indels 60; Gaps 11;

QY 36 ROPGVCHGTACCCYGRNRNKGVCATCEPG-----CKRGE 73
 DB 142 RNPILCRGCTC-----NTEGSRCDCCPGHHISPNISACIDINECDLSTNLCRNH 193
 QY 74 CV---GPNKRCFPGY---TGKTCSDYNECGMKPRPCOHRCVNTHGSKFCFLSGHML 127
 DB 194 CVNLIGKQACNPGYQSTADKLHIDIDECGINNGGCENFTGSEGSCKGKGFALM 253
 QY 128 PDATCVNSRTGAMINCOYSCEDTE---EGPQ-----CLPSSGLRLAPNGRDCLDI 175
 DB 254 PD-----HRTCTDID---ECEDPNPICDGGCTNIPGEYRCLC-YDGFMASEDMDKTCVDY 304
 QY 176 DECAAGKVICPYNRCVMTFGSYCKCHIGFELQYISGRYDCIDINECTMDSHTCSSHAN 235
 DB 305 NECDLHPNIC-LSGTCENKTSFICHCDMGYSK--KGTGCTDINECEIGHANCDRHAV 361
 QY 236 CFNTGSEFKCKCKGKGYKGLRCSAIPENS 265
 DB 362 CTNIPGSFKCSOSSGWMINGIKCTDLDECS 391

RESULT 12
 Q9WUH8 PRELIMINARY; PRT; 2872 AA.
 AC Q9WUH8:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FIBRILIN-1.
 GN FBN1.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A.

RX MEDLINE=99032689; PubMed=9815129;
 RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,
 RA Peterson D.R.,
 RT "Isolation of rat fibrillin-1 cDNA and its relevance in metanephric
 development."
 RT Am. J. Physiol. 275:F710-F723(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kanwar Y.S.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF135059; AAD34438.1; .
 DR HSSP; P35555; 1APJ.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR InterPro: IPR000822; ZnF-C2H2.
 DR Pfam; PF00008; EGF; 45.
 DR Pfam; PF00683; TB; 9.
 DR SMART; SM00179; EGF_CA; 41.
 DR SMART; SM00001; EGF_Like; 5.
 DR PROSITE; PS00010; ASX_HYDROXYL; 42.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; 38.
 DR PROSITE; PS01187; EGF_CA; 41.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 2872 AA; 312069 MW; 0C4F7F3B87A80280 CRC64;

Query Match 12.6%; Score 385.5; DB 11; Length 2872;
 Best Local Similarity 28.2%; Pred. No. 3.7e-22;
 Matches 95; Conservative 37; Mismatches 92; Indels 113; Gaps 13;

QY 38 PGVCHGT-----KLACCYGM-----RNSKGVCAATCEPG-CKFGEV---GPNKC 80
 DB 1037 PSLCTHYGKCRNTIGSEFKCSDGFAIDSEERCTDIDECRISPDLGCRQCVNTPGDFEC 1096
 QY 81 RCFPGYTG-----KTCSDVNECGMKPRPCOHR-CVNTGSKYKFCFLSGHMLPMDT 131
 DB 1097 KCDEGESEGFMMKNC-MDIDECQRPDLRCRGICHTNGSTRCECPSHQSLSPNISACI 1155
 QY 132 -----CNSRTGAMIN---COYSCED 149
 DB 1156 DINECELSANLCPHGRCVNLICKYBECANPGYHPHTRDLFCVYDIDECISLNGGCEFCFN 1215
 QY 150 TEEGQCCLPSSGLRLAPNGRDCLDIDECASGVIC-----PYNRVCV----- 192
 DB 1216 SDGSYECSC-QGFALMPORSCCTDIDECEDPNICDGGCTNIPGEYRCLCYDGFMAE 1274
 QY 193 -----NTEGSYCKCHIGFELQYISGRYDCIDINECTMDSH 228
 DB 1275 DMKTCVDVNECDLNPNICLSGTCENKTSFICHCDMGYSK--KGTGCTDINECEIGHAN 1332
 QY 229 TCSHANCENFTGSEFKCKCKGKGYKGLRCSAIPENS 265
 DB 1333 NCGRHAVCTNTAGSFKCSGPMWIGDKICXTDLDECS 1369

RESULT 13
 O88840 PRELIMINARY; PRT; 3857 AA.
 AC O88840:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MUTANT FIBRILIN-1.
 GN FBN1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.

RC STRAIN-B10.D2; PubMed=9405934;
 RX BOLAINE=98069008; PubMed=9405934;
 RA Bora C.A., Murai C., Casares S., Kasturi K., Nishimura H., Honjo T.,
 Matsuda F.;
 RT "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)
 mouse.";
 RL DNA Res. 4:267-271(1997).
 DR EMBL: AF007248; AAC62317.1; -
 DR HSSP: P35555; IAPJ.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR002557; Cnltin binding.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF00008; EGF; 64.
 DR Pfam: PF00683; TB; 12.
 DR SMART: SM00494; Chbd2; 2.
 DR SMART: SM00179; EGF_CA; 60.
 DR SMART: SM00001; EGF_like; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 61.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE: PS01186; EGF_2; 50.
 DR PROSITE: PS01187; EGF_CA; 61.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KW Calcutin-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 3857 AA; 418301 MW; 5BC0618BC527E04C CRC64;

Query Match 12.5%; Score 383.5; DB 11; Length 3857;
 Best Local Similarity 32.2%; Pred. No. 7.9e-22;
 Matches 86; Conservative 34; Mismatches 82; Indels 65; Gaps 11;

QY 39 GVCHYGLTLACCYGWRNRSGVCEATCEPG-----CKRGEV- 75
 Db 1129 GICH-----NTEGSRCECPRGHQLSPNISACIDINCELSANICPHGRCN 1175
 QY 76 --GPKKCRFPY---TGKTSQDVNEGCMKRPCOHRVNTGSGYKCCLSGHMIMPA 130
 Db 1176 LIGKQACACNPGYHPTDHLRFLCVDIDECISIMNGCETCTNSDSEYSCQCPGFALMPD- 1234
 QY 131 TCVNSRTCAINCOYSCEDTE---EGPO-----CLCPSSGLRLAPNGRCDLIDEC 178
 Db 1235 ----QRCTDID---ECEDNPICDGGCCTNIPGEYKCLC-IDGFMASDMMTCVDVNC 1286
 QY 179 ASGVICYEYNRRCVNTFSGSYKCHIGFELYISGRYDCIDINECTMDSHTCSHANCEN 238
 Db 1287 DLNPMIC-LSGTCENTKSEFICDGMYSK--KKGTCGTIDINECEIGHANCGRHAVCTN 1343
 QY 239 TQGSFKCKCKQGYKNGRLCSAIPENS 265
 Db 1344 TAGSFKSCSPGMIDGAIKCTDIDEC 1370

RESULT 14
 Q99K58 PRELIMINARY; PRT; 1174 AA.
 AC Q99K58;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SIMILAR TO FIBULIN 2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMARY TUMOR. MAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
 RC TISSUE:;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC005443; AAH05443.1; -
 DR HSSP: P00736; IAPQ.

DR InterPro: IPR000020; Anaphylatoxin.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR Pfam: PF01821; ANATO; 2.
 DR Pfam: PF00008; EGF; 6.
 DR SMART: SM00104; ANATO; 3.
 DR SMART: SM00179; EGF_CA; 11.
 DR SMART: SM00179; EGF_CA; 9.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 3.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 3.
 DR PROSITE: PS00010; ASX_HYDROXYL; 5.
 DR PROSITE: PS01186; EGF_2; 5.
 DR PROSITE: PS01187; EGF_CA; 9.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 KW Calcutin-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 1174 AA; 126460 MW; 8D628AC710FBAB8 CRC64;

Query Match 12.5%; Score 382; DB 11; Length 1174;
 Best Local Similarity 28.5%; Pred. No. 2.3e-22;
 Matches 114; Conservative 39; Mismatches 139; Indels 108; Gaps 20;

QY 47 LACCYGWRNRSG---VCEATCEPG---CKRGE-CV---GPKKCRFPGYT-----GKTC 92
 Db 831 LVCGRGYHANEGSSECVYDNECEFTGYHRCGEQCLYNLPGSTRCCKPFGQDAREGTCT 890
 QY 93 ODVNEGCMK-RPCOHRVNTGSGYKCCLSGHMIMPAT-CVNSRTCAINCOYSCEDT 150
 Db 891 -DVNECWSPGRICQHTCENTPGSYRSCAAGFLAADGKHCEDEVNECETRSCDECA 949
 QY 151 EGGPQCLPSSGLRLAPNGRCDLIDECASGR-VICPYNRKRVNTFSGYKCHIGFELD 209
 Db 950 YSYCYCYC-RQGYQLADEHTDCTDIDECAGAGILCTF--RCVNVPGSYQACPEQGYTM 1006
 QY 210 YISGRYDCIDINECTMDSHTCSHANCENFTQGSFC---CKQGY-KGNGLCSAIPENS 265
 Db 1007 MANGR-SCKDIDECALGTHNCSEAECHNIOGFRCLRDCEPNRYVSETGTCERTCD 1065
 QY 266 VKEVLRAPGTIDRIKLLAHNNSMKKAKIKNVPEPTPTPKVNILOPNYEELVSRG 325
 Db 1066 ITECOTSPARI---THYQNFQGLVPAHIFRIGFAP----- 1100
 QY 326 GNSHGKKNEMKMEGLEDEKREKALKNDIEBSLRGDFV---FPKYNEAGEF----- 377
 Db 1101 -----AFADGTLSLTITTKNGEGYFVTRRL 1125
 QY 378 ---GLIVQRKALSKLEHKDINISVDCSFNHGICDMKO 413
 Db 1126 NAYTGVSLSR---SVLEPRDFALDVEKML-----WRQ 1155

RESULT 15
 Q9BI05 PRELIMINARY; PRT; 2189 AA.
 AC Q9BI05;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MICRONEME PROTEIN 4.
 GN MIC4.
 OS Eimeria tenella.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
 OC Eimeria.
 NCBI_TaxID=5802;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HOUGHTON;
 RA Tonley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.;
 RT "Etic4: A microneme protein from Eimeria tenella that contains tandem
 RT arrays of epidermal growth factor-like and thrombospondin type-1
 RT repeats.";
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ306453; CAC34726.1; -
 DR HSSP: P35555; IEMN.

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00008; EGF_24.
DR SMART: SM00181; EGF_30.
DR SMART: SM00179; EGF_CA_30.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS01186; EGF_2; 18.
DR PROSITE: PS01187; EGF_CA; 18.
DR PROSITE: PS50092; TSPI; 1.
KW Calcium-binding; EGF-like domain; glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2189 AA; 224388 MW; C272A420B94FCB2D CRC64;

Query Match 12.2%; Score 374.5; DB 5; Length 2189;
Best Local Similarity 32.2%; Pred. No. 2.1e-21;
Matches 93; Conservative 32; Mismatches 77; Indels 87; Gaps 14;
QY 44 GTKLAC-CY-GMRRNSKGVCEATCEPGCK-----FGE---CV---GPNKCRC 82
DB 966 GTEAICTCHSGYEGNGEG-----EEGCKNIDECVSGEPCKDFGEGYCVDSFGSFGSC 1019
QY 83 PPGYTGKTCG-QDVNEC-GMKPRPCQ--HRCVNTGSGYKCFCLSG----- 123
DB 1020 ATGFIRKRCQDIDECIDGCKMNTCAPYGICTNTVGSFTCSAAGFTGDGILCEDIDEC 1079
QY 124 ----HMLMPDATCYNSRTCAMINQ-----YCEDTEEGPQCLCPG----- 160
DB 1080 ATAAHTCDPNATYCVNTVGSFECGCKEGFSGDGHTCTDIDE--CADPMLNKCDTHKGIQ 1136
QY 161 -----SGLRLAPNGRDCIDIDECASGKVICPYNRRCVNTFGSYCKCHIGFELQ 209
DB 1137 NGTSGYTCGCGRPGYSLADGFTCDNVDECAAGTATCGERSFCVDIQSGYKCECKNG----- 1192
QY 210 YISGRYDCIDINECTMDSHTCSHANCFNTQGSFCKCKGQYKGNGLRC 258
DB 1193 YRQCGEDCVVDYDECEADVHTCSEHATCTNTEGSHTCCTCNEGYQGDGKRC 1241

Search completed: October 17, 2002, 16:01:27
Job time : 40 secs

THIS PAGE BLANK (USPTO)